Lost Sales and Emergency Order Systems under Stuttering Poisson Demand

by Jie Chen

This thesis/dissertation document has been electronically approved by the following individuals:

Jackson,Peter (Chairperson)

Durrett,Richard Timothy (Minor Member)

Henderson,Shane G. (Minor Member)

LOST SALES AND EMERGENCY ORDER SYSTEMS UNDER STUTTERING POISSON DEMAND

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Jie Chen

August 2010

LOST SALES AND EMERGENCY ORDER SYSTEMS UNDER STUTTERING POISSON DEMAND

Jie Chen, Ph.D.

Cornell University 2010

We investigate the (S-1,S) inventory policy under stuttering Poisson demand and exponentially-distributed lead times when demand in excess of on-hand inventory is routed to an emergency order fulfillment system. This system contains a regional stocking location (RSL), which serves two types of facilities: a set of field service locations (FSL) and an emergency stocking location (ESL). The field service locations support technical service representatives who make visits to customer sites to repair equipment. We derive both exact and approximate expressions for the mean and variance of the number of units in emergency resupply. We also estimate the probability of zero units in emergency resupply. Simulation results confirm the quality of these approximations. Later, we use a distribution with an atom at zero and a zero-truncated negative binomial distribution to approximate the shape of this distribution. The quantiles are shown to be well approximated in simulations with various settings. In particular, the approximation is excellent in the upper tail which is the portion of the distribution used to determine the target inventory level for the emergency stocking location. Finally, we develop an optimization algorithm for setting stock levels in such a system with both field service locations and an emergency stocking location. The problem is an integer programming problem with a potential non-convex objective and we explore a heuristic algorithm for solving the optimization problem. For empirical studies, we compare the results of our

heuristics with PSWARM, a general purpose algorithm for such problems.

BIOGRAPHICAL SKETCH

Jie Chen was born on September 18, 1983 in Chongqing, China. She attended middle school and high school at Bashu Middle School, Chongqing, China.

In 2001, she joined the School of Mathematical Sciences at Peking University and earned a B.S. in Statistics in June of 2005.

In August of 2005, she entered the School of Operations Research and Information Engineering, Cornell University. She received a M.S. in 2008 and completed her Ph.D. in Applied Operations Research in June 2010. Her research interests are inventory control, supply chain management, stochastic modeling, and simulations.

This thesis is dedicated to my family.

ACKNOWLEDGEMENTS

I would like to thank all people who have helped and inspired me during my doctoral study at Cornell.

First of all, I would like to heartily thank my Ph.D. major supervisor, Prof. Peter L. Jackson for his passionate encouragement, patient guidance and excellent support during my doctoral studies. He led me into this wonderful world of supply chain management area and always provided instructive guidance, sound advice and great ideas. Because of his kindness, enthusiasm and expertise, my research life was smooth, inspiring and rewarding. Furthermore, his impact on me is never limited to the research. In the early days, he spent extra hours reading a heroic fantasy novel "Watership Down" with me, which greatly improved my pronunciation and rhythm. As being one of his teaching assistants, his passion in teaching and meticulousness impressed me and cultivated my positive work attitudes and values. I benefited greatly from all the experiences.

My sincere gratitude also goes to Prof. John A. Muckstadt for his advice, inspiration, impressive rigor, enormous pool of knowledge and crucial contribution to my Ph.D. study and research. He was always accessible and willing to help me with any problem I encountered. His support with his professional insight, rich experience and original creativity enlightened me and nourished my capacity and intellectual skills. I am very grateful for my years working together with Prof. Jackson and Prof. Muckstadt.

Besides, I would like to thank the rest of my thesis committee: Prof. Shane G. Henderson and Prof. Richard T. Durrett for their encouragement, insightful comments, and challenging questions. I would also like to thank Prof. Gennady Samorodnitsky for contributions to the approach taken in sections 3.3 and 3.4.3.

I am grateful for valuable discussions with Prof. Sidney I. Resnick and Prof. David B. Shmoys. I am grateful to Prof. Phillip E. Protter for his enthusiastic teaching and supportive advice on my study. I am indebted to all the professors and teachers who taught me and influenced my academic interests in the past 21 years. I am grateful to graduate thesis advisor for providing helpful comments on the manuscript.

I would also wish to thank Ms. Elizabeth Appleton, Ms. Selene Cammer, Mr. Jonathan Froehlich, Mr. Pat Gillespie, Ms. Cindy Jay, Mr. Eric Johnson, Ms. Kathryn M. King and Mr. David Schneider for helping me with my work at the School of Operations Research and Information Engineering. Special thanks are also due to my friends and colleagues.

The final and deepest thanks go to my family, in particular to my parents and my husband, for all their unreserved support and encouragement over the years. This thesis is dedicated to them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Resupply networks for service parts are large scale, complex systems that must achieve high performance in terms of both short resupply times and low system cost. Hundreds of stocking locations and hundreds of thousands of stock keeping units are managed within many of these systems. Typically there are several supply echelons, and lead times between echelons vary considerably. There are also emergency stocking locations and many different ways to resupply individual stock locations. Additionally, there is considerable variation in demand over time at the lowest echelon of the system. Optimizing stock levels at all the locations in such systems is challenging and interesting.

The particular environment we consider in our research is that faced by a large company that provides service (repair and maintenance) to reprographic equipment at customer sites throughout the United States. Customer service technicians operate from the field service locations (FSL) to visit customer sites and perform repairs or maintenance. These FSLs maintain small amounts of inventory of the most commonly needed parts. The demand for these parts by the customer service technicians exhibits a high variance-to-mean ratio. When a service part is withdrawn from inventory by a technician, a replenishment order is placed immediately with the regional stocking location (RSL). There are only a handful of RSLs to serve all the replenishment demand for the county. From the time an FSL places an order until it receives it from the nearest RSL is approximately two weeks. If a technician requires a service part and the field service location does not have the part in stock then an emergency order is placed to the nearest emergency stocking location (ESL). This can occur if the FSL has chosen

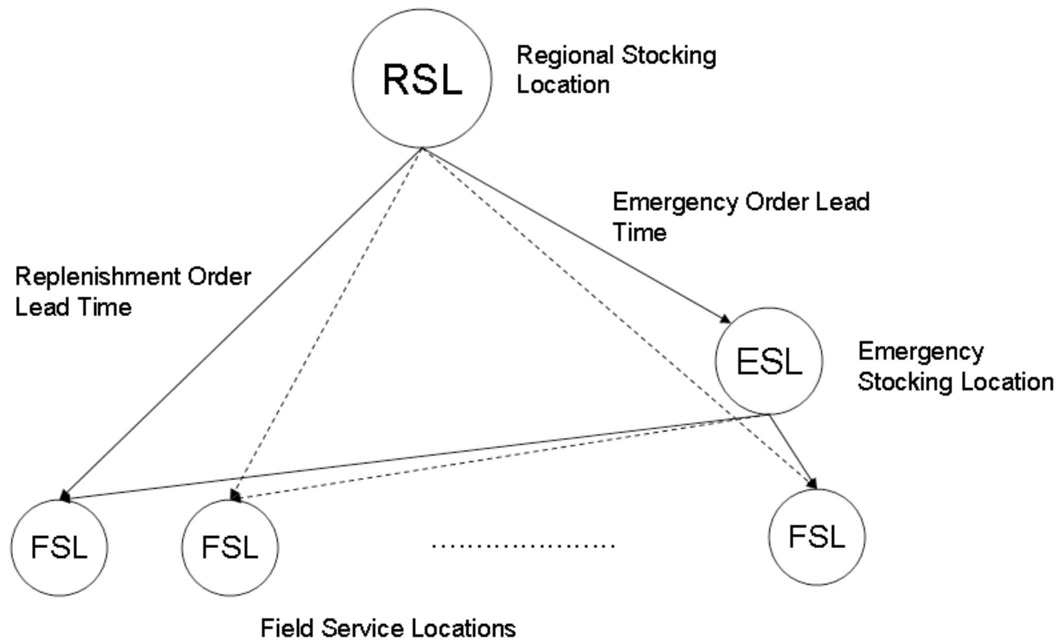

Figure 1.1: A System with Emergency Resupply

not to stock this part at all because of low demand rates; but, it can also occur because previous demands have exhausted the on hand inventory at the FSL and the replenishment orders are still outstanding. There are numerous ESLs distributed throughout the country. The time for an FSL to receive an emergency order, if the closest ESL has stock of the part on hand, is quite short. Often the part can be delivered on the same day of business that it was ordered. When an emergency order is placed with an ESL, a replenishment order is placed immediately with the nearest RSL. The leadtime to fulfill these emergency replenishment orders is approximately two weeks, the same amount of time as required for regular replenishment orders to FSLs. The emergency stocking network can be partitioned into clusters of the FSLs served by their nearest ESL. Figure 1.1 shows the basic resupply network for a single cluster of FSLs.

This is an efficient system from an inventory viewpoint because the demand for parts has a high variance-to-mean ratio at the FSL but these demands can be pooled at the ESL. This can reduce total safety stock requirements without sacrificing customer service. The challenge is to predict customer service levels accurately as a function of stocking policies at the ESL and the associated cluster of FSLs, and to optimize these stocking policies for a given inventory budget.

We make a number of simplifying assumptions to permit an analytical approach to this challenge. First, all the FSLs and the ESL employ an (S-1,S) inventory policy. At the FSLs, the demand arrival process is hypothesized to be a stuttering Poisson demand process. The stuttering Poisson process is a special case of a compound Poisson process in which the compounding order size distribution is geometric. This assumption allows us to capture the the high variance-to-mean ratio attribute of the demand better than a simple Poisson arrival assumption. We also assume that the demand process at each FSL is independent of those at other FSLs. Next, we assume that the distribution of the lead time from the RSL to the FSLs or the ESL is exponential. This allows us to analyze the system using continuous time Markov processes. We later show that this assumption is not critical. Our results could be applied to generally distributed lead times and, empirically, the model performs well in comparison to simulation. However, even with these assumptions this system is still not easy to solve due to the non-constant order size of the demand. This requires formulating the state space to keep track of the order sizes in resupply; the size of the state space grows quickly as the target stock levels at the FSLs increase.

Our response to the overall challenge of prediction and optimization is presented in the form of three sequential papers that progressively address these issues. In the first paper we develop a predictive model for a single FSL by understanding the emergency orders to have the same behavior as in a lost sales model. In the second paper we develop a predictive model for the ESL by extending the results of the first paper. In the third paper we develop an algorithm to optimize stock levels at the ESL and across the associated cluster of FSLs simultaneously.

The specific contributions of each paper are as follows.

In chapter 2, we consider the emergency orders to be lost sales and focus on the number of units on order at one FSL. We investigate the (S-1,S) inventory policy under stuttering Poisson demand and generally distributed lead times. First, the theory of continuous reversible Markov chains is applied to the model when the lead time is exponentially distributed. Later, we prove that the formula derived is insensitive to the lead time distribution. At the same time, we note an error in Feeney and Sherbrooke's paper (1966), which is a seminal paper deriving an analytical formula for the lost sales case under compound Poisson demand. We demonstrate empirically that their error does not greatly affect the optimal stock levels. We claim that exact analysis for the general lost sales distribution is still an open question since our method applies only to the case of stuttering Poisson demand. We also observe that as the variance-to-mean ratio of the demand process increases, the number of outstanding units on order will arrive more closely to each other. This phenomenon can be expressed as "orders tend to arrive later when you are out of stock".

In chapter 3, we add a single ESL to the previous model and decompose the state space for the emergency order system to analyze the mean and variance of the number of units in emergency resupply. The expressions we develop can be used to obtain their values exactly; however, we have constructed an approximation method to make calculating their values much simpler. We also construct a method for estimating the probability of zero units in emergency resupply. Simulations show that the approximations are quite accurate. Using these methods to estimate parameters, we approximate the steady state distribution of the number of emergency ordered units using an atom at zero and a zero-truncated negative binomial distribution. Simulation results confirm the quality of this approximation in the upper tail, which is the main portion of the distribution used to determine the target inventory for the ESL.

In chapter 4, we use the steady state distributions of the number of units in resupply (regular replenishment and emergency resupply) to determine the best stock levels at the FSLs and ESL across all part numbers to minimize backorder and emergency resupply costs subject to a network budget constraint. The problem is an integer programming problem with a potential non-convex objective and we explore a heuristic algorithm for solving the optimization problem.

Concluding comments and suggestions for further research are found in chapter 5.

BIBLIOGRAPHY

[1] Feeney, G.J., Sherbrooke, C.C., The (S-1,S) Inventory Policy under Compound Poisson Demand, *Management Science,* Vol. 12, No. 5, Series A, Sciences., pp. 391-411 (Jan.,1966).

CHAPTER 2

EXACT ANALYSIS OF A LOST SALES MODEL UNDER STUTTERING POISSON DEMAND

Jie Chen, Peter L. Jackson, John A. Muckstadt
School of Operations Research and Information Engineering,
Cornell University, Ithaca, NY, USA
jc562@cornell.edu, pj16@cornell.edu, jam61@cornell.edu

Abstract: We investigate the (S-1,S) inventory policy under stuttering Poisson demand and generally distributed lead time when the excess demand is lost. We correct results presented in Feeney and Sherbrooke's seminal paper (1966). We also prove that the distribution of "ordered unit delivery times" becomes increasingly concentrated as the variance-to-mean ratio of demand increases.

2.1 Introduction

In a seminal paper, Feeney and Sherbrooke (1966) model and analyze the (S-1,S) inventory policy under lost sales, compound Poisson demand and general lead time distributions. They consider two variants of lost sales: orders arriving with insufficient stock on hand to be completely filled may be either completely lost (the complete fill case) or they may be partially filled with remaining stock on hand with only the balance of the order lost (the partial fill case). Feeney and Sherbrooke derive the stationary distribution of the number of units in resupply. Their result extends Palm's theorem to the case of compound Poisson demand.

Feeney and Sherbrooke initially pose the model in terms of a state space that tracks all outstanding orders and their order sizes but conduct their analysis using a simpler state space that considers only the number of orders and number of units outstanding.

In an unpublished paper, Baganha (1985) notes an inconsistency in the proofs of the Feeney and Sherbrooke results but arrives at the same results. Apart from this, the Feeney and Sherbrooke results have been unchallenged.

By focusing on the case when the compounding order size distribution is geometric (the so-called "stuttering Poisson demand process"), we find that the model using the original state space identified by Feeney and Sherbrooke is amenable to analysis. We show that it possesses the property of reversibility and we use this property to derive the stationary distribution of the number of units outstanding. The result is established initially for exponentially distributed lead times but then extended to general lead time distributions. The property of reversibility, and hence the result, is shown to be peculiar to the geometric distribution. The result, for the partial fill case, does not agree with Feeney and Sherbrooke's result and, though it agrees with their result for the complete fill case, it does not satisfy their version of the steady state conditions. We conclude that the Feeney and Sherbrooke results are not exact. The error can be traced to their assumption that the order sizes in resupply have the same distribution as the order sizes of arriving orders. Since orders are filtered by the lost sales mechanism, this assumption does not hold in general. The stationary distribution of the number of units outstanding for general compound Poisson demand is, therefore, still an open question. Later, we also demonstrate that, at least for the stuttering Poisson case, the F-S formulas are good approximations when used to set optimal stock levels. We also prove an interesting result on the spread of expected order replenishment delivery times as a function of the variance-to-mean ratio of the demand process. We show that the spread of these times decreases and becomes more concentrated as the variance-to-mean ratio grows.

This is not the first paper to apply reversibility in the context of an inventory model. Moinzadeh (1989) considers a variant of the (S-1,S) inventory model under Poisson demand in which an order is lost with probability α_j if there are j backorders at the time of its arrival, and backordered otherwise. He derives the stationary distribution for the case of exponentially distributed lead times and conjectures that the result would extend to generally distributed lead times. Smeitink (1990) uses the concept of quasi-reversibility to establish the more general result.

Our paper is organized as follows. In section 2.2, we introduce the lost sales model and notation. Our focus is on the partial fill case in that section. In section 2.3, we construct a Markov chain to represent state transitions and show the reversibility of this Markov chain. Using this property in section 2.4, we derive the stationary distribution of the number of units on order for the lost sales model. In section 2.5, we compare these exact results with Feeney and Sherbrooke's results. In section 2.6, we prove a property of the expected ordered unit delivery times. Concluding comments are found in section 2.7. A consideration of the complete fill case, and proofs for all theorems are found in the appendix.

2.2 A Lost Sales Model with Compound Poisson Demand and Exponential Lead Times

We consider a continuous time model in which demand arrives according to a stationary compound Poisson process. Let λ denote the rate of arrivals of customer orders and let X denote the order size, which is a positive, integer-valued, random variable. Let $p_k \equiv P\{X = k\}$ and let $\overline{P}_k \equiv P\{X > k\}$ for all $k = 0, 1, 2, ...$ We assume at least one unit is ordered for each customer arrival: $p_0 = 0$ and $\overline{P}_0 = 1$, although the results are easily generalized to allow for zero-sized orders. For the special case of the so-called stuttering Poisson process, the order size distribution is geometric. Let p denote the probability of a unit-sized order under the geometric distribution: $p_1 = p$. In this case, for all $k = 1, 2, ...,$ $p_k = p(1-p)^{k-1}$ and $\overline{P}_k = (1-p)^k$.

Let I_t denote the inventory on hand at time t, $t \geq 0$, a non-negative, integer-valued random variable. We assume that demand in excess of inventory on hand is lost but that a customer's order may be partially filled.

We assume initially that lead times for replenishment orders are independent, exponentially-distributed random variables with rate μ. This assumption is relaxed later. Let τ denote the expected replenishment order lead time: $\tau = 1/\mu$. We assume that the system is managed according to an $(S-1, S)$ policy.

Let N_{kt} denote the number of replenishment orders of size k outstanding at time t, for $k = 1, 2, ..., S$, and let $N_t = (N_{kt})_{k=1}^{S}$ denote the vector of outstanding replenishment orders. Given our assumptions of lost sales, partial fills, and an

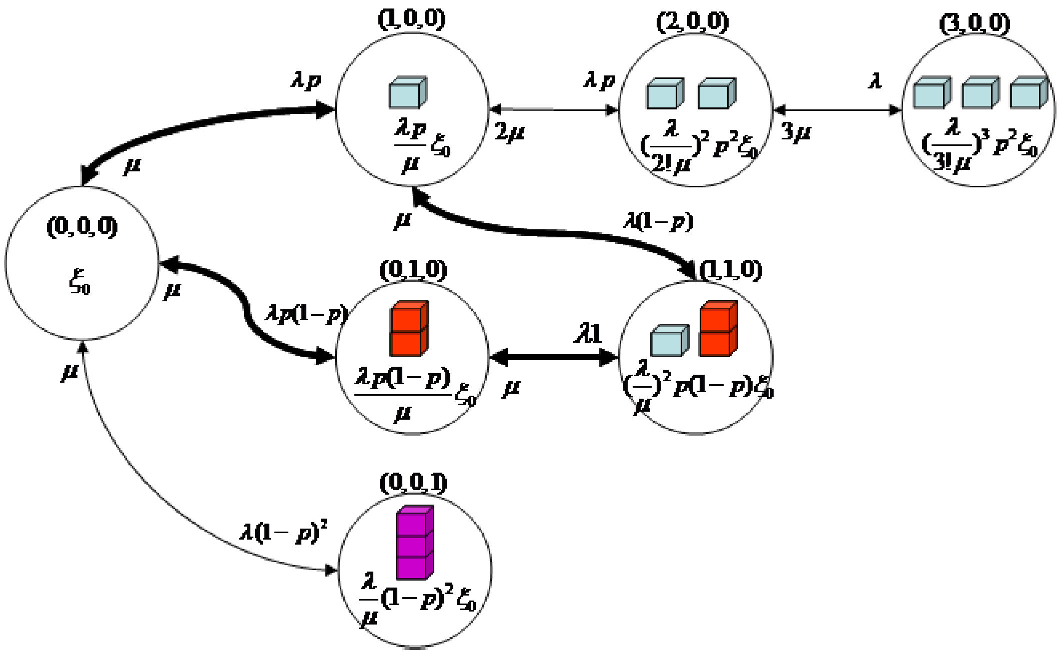

Figure 2.1: State Space and Single Order Transitions for $S = 3$ and Partial Fill Case

$(S-1, S)$ policy, it follows that

$$I_t + \sum_{k=1}^{S} kN_{kt} = S.$$

The stochastic process $N = \{N_t, t \geq 0\}$ is a finite-state, time-homogeneous Markov process. Let V index the state space of the underlying Markov chain. For each $i \in V$, we denote the vector of outstanding replenishment orders by $n(i) = (n_1(i), n_2(i), ..., n_S(i))$, where $n_k(i) \in \{0, 1, ..., \lfloor S/k \rfloor\}$ forall$k = 1, ..., S$, and $\sum_{k=1}^{S} kn_k(i) \leq S$. Furthermore, the implied number of units on hand is given by

$$n_0(i) \equiv S - \sum_{k=1}^{S} kn_k(i).$$

Let $m(i) \equiv \sum_{k=1}^{S} n_k(i)$, be the total number of outstanding orders in state i. V is the state space initially considered by Feeney and Sherbrooke.

The graphic in Figure 2.1 illustrates the possible states when $S = 3$.

Let the pair (i, j) denote a transition from state $i \in V$ to state $j \in V$. Let

$$\|(i, j)\| \equiv \sum_{k=1}^{S} |n_k(i) - n_k(j)|\,, \text{forall}(i, j) \in V \times V,$$

the number changes in outstanding order levels separating i from j. State transitions occur only when either a customer order arrives or a replenishment order arrives. Since the probability that two or more orders arrive simultaneously is zero, we focus on single-order transitions, that is, transitions for which $\|(i, j)\| = 1$. The arrows in Figure 2.1 indicate all possible single order transitions for the $S = 3$ case. For a single-order transition (i, j), let k_{ij} denote the size of the (accepted) customer order or the size of the arriving replenishment order, as appropriate:

$$k_{ij} \equiv \sum_{k=1}^{S} k\,|n_k(i) - n_k(j)|\,\text{forall}(i, j) \in V \times V \text{s.t.}\,\|(i, j)\| = 1.$$

We classify single-order transitions by whether they are customer order arrivals ($(i, j) \in V_C^2$) or replenishment order arrivals ($(i, j) \in V_R^2$):

$$(i, j) \in \begin{cases} V_C^2 & \text{iff} \quad \|(i, j)\| = 1 \text{ and } n_{k_{ij}}(i) < n_{k_{ij}}(j) = n_{k_{ij}}(i) + 1 \\ V_R^2 & \text{iff} \quad \|(i, j)\| = 1 \text{ and } n_{k_{ij}}(i) > n_{k_{ij}}(j) = n_{k_{ij}}(i) - 1 \end{cases}$$

It is easily seen that the infinitesimal generator for this Markov process N is given by

$$A_{ij} \equiv \begin{cases} n_{k_{ij}}(i)\mu & \text{if}(i, j) \in V_R^2, \\ \lambda p_{k_{ij}} & \text{if}(i, j) \in V_C^2, n_0(j) > 0, \\ \lambda \overline{P}_{k_{ij}-1} & \text{if}(i, j) \in V_C^2, n_0(j) = 0, \\ -(m(i)\mu + \lambda 1\{n_0(i) \neq 0\}) & \text{if}\, j = i, \\ 0 & \text{otherwise}, \end{cases}$$

where $1\{E\}$ is the indicator function of condition E (i.e. $1\{E\} = 1$ if E and $= 0$ otherwise). To see this, note that $(i, j) \in V_R^2$ means a replenishment of size k_{ij}

arrived. In this case, the transition rate is $n_{k_{ij}}(i)\mu$. The condition $(i, j) \in V_C^2, n_0(j) > 0$ means a new customer order of size k_{ij} arrives and could be satisfied. So the transition rate is $\lambda p_{k_{ij}}$. The condition $(i, j) \in V_C^2, n_0(j) = 0$ means an order arrives with order size greater than or equal to the on-hand inventory level causing $n_0(j) = 0$ with transition rate $\lambda \overline{P}_{k_{ij}-1}$. For any other (i, j), $j \neq i$, there is no single step transition between them, so the transition rate is zero. Finally, for $j = i$:

$$A_{ii} = -\sum_{j \in V, j \neq i} A_{ij} = -(\sum_j n_{k_{ij}}(i)\mu + \lambda 1\{n_0(i) \neq 0\}) = -(m(i)\mu + \lambda 1\{n_0(i) \neq 0\}).$$

In the case of the stuttering Poisson demand process, this infinitesimal generator simplifies to:

$$A_{ij} = \begin{cases} n_{k_{ij}}(i)\mu & \text{if}(i, j) \in V_R^2, \\ \lambda p^{1\{n_0(j)>0\}} (1-p)^{k_{ij}-1} & \text{if}(i, j) \in V_C^2, \\ -(m(i)\mu + \lambda 1\{n_0(i) \neq 0\}) & \text{if}\, j = i, \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

Important properties of this infinitesimal generator will be seen to hold only if the demand process is a stuttering Poisson process.

2.3 Reversibility

A reversible continuous-time Markov chain is defined and described in Resnick (2005, p433-434). The following proposition characterizes the reversible property.

Proposition 1 *A stationary Markov chain* $\{\tilde{X}(t), -\infty < t < \infty\}$ *is reversible if and only if when* $\tilde{A}$ *is the generator matrix of* $\left\{\tilde{X}(t)\right\}$, *the detailed balance equations*

$$\tilde{\xi}_i \tilde{A}_{ij} = \tilde{\xi}_j \tilde{A}_{ji}, \text{forall} i \neq j, \tag{2.2}$$

hold for some probability distribution $\tilde{\xi}$. If a probability distribution $\tilde{\xi}$ can be found to (2.2), then $\tilde{\xi}$ is, in fact, the stationary distribution of $\left\{\tilde{X}(t)\right\}$.

For the replenishment order process, N, we choose as a reference state the state i_0 for which no orders are outstanding ($n_0(i_0) = S$). For any state $i \in V$, with at least one outstanding order ($n_0(i) < S$), we seek to define a sequence of single-order transitions that will lead from i to the reference state, i_0. It is natural to choose each transition to correspond to the delivery of a replenishment order. In this case, the number of transitions required will be given by the total number of outstanding orders in state i. We form a path of states $j_0 = i \rightarrow j_1 \rightarrow j_2 \rightarrow ... \rightarrow j_{m(i)-1} \rightarrow j_{m(i)} = i_0$ in which each transition (j_l, j_{l+1}) corresponds to the delivery of a replenishment order ($(j_l, j_{l+1}) \in V_R^2$). Furthermore, for each transition, we choose the size of the arriving replenishment order according to a *largest subscript rule*. That is, let j_l denote a state on this path, $l = 0, 1, ..., m(i) - 1$. Let k_l denote the order size of the largest outstanding order: $k_l = \max\{k \in \{1, 2, ..., S\} : n_k(j_l) > 0\}$. We choose as the next state, j_{l+1}, the state corresponding to the arrival of a replenishment order of size k_l. That is, j_{l+1} is the unique state satisfying $n_k(j_{l+1}) = n_k(j_l) - 1\{k = k_l\}$ for all $k = 1, ..., S$. It should be clear that a path of single-order transitions from the reference state i_0 back to state i can be found by simply reversing the sequence: $i_0 \rightarrow j_{m(i)-1} \rightarrow ... \rightarrow j_1 \rightarrow i$. Along this reverse path, the transitions all correspond to customer arrivals ($(j_l, j_{l-1}) \in V_C^2$).

Suppose that the replenishment order process is reversible and that η is the stationary distribution. Given the largest subscript rule of selecting paths between any state $i \in V$ and the reference state i_0, observe that repeated application of (2.2) yields the following:

$$\xi_i A_{ij_1} A_{j_1 j_2} ... A_{j_{m(i)-1} i_0} = \xi_{i_0} A_{i_0 j_{m(i)-1}} A_{j_{m(i)-1} j_{m(i)-2}} ... A_{j_1 i}.$$

This suggests a solution of the form $(i \in V)$:

$$\xi_i = \begin{cases} \nu_i \xi_{i_0} & i \neq i_0, \\ \frac{1}{1+\sum_{j \neq i_0} \nu_j} & i = i_0, \end{cases} \tag{2.3}$$

where

$$\nu_i \equiv \frac{A_{i_0 j_{m(i)-1}} A_{j_{m(i)-1} j_{m(i)-2}} ... A_{j_1 i}}{A_{ij_1} A_{j_1 j_2} ... A_{j_{m(i)-1} i_0}}. \tag{2.4}$$

In Figure 2.1, for the $S = 3$ example, arrival transition rates are shown above the transition arcs while delivery transition rates are shown below the transition arcs. The path (0, 0, 0)-(1, 0, 0)-(1, 1, 0) illustrates the longest subscript rule. Vaules for ξ_i are shown within each node. For $n(i) = (1, 1, 0)$, (2.4) yields $\nu_i = \frac{\lambda p \cdot \lambda(1-p)}{\mu \cdot \mu} = \frac{\lambda^2 p(1-p)}{\mu^2}$, and (2.3) yields $\xi_i = \frac{\lambda^2 p(1-p)}{\mu^2} \xi_{i_0}$ as shown.

Proposition 2 *For the geometric order size distribution, the suggested solution (2.4) is given by* $(i \in V)$:

$$\nu_i \equiv \frac{\left(\frac{\lambda p}{\mu(1-p)}\right)^{m(i)}}{\prod_{k=1}^{S} (n_k(i)!)} \frac{(1-p)^{S-n_0(i)}}{p^{1\{n_0(i)=0\}}}. \tag{2.5}$$

Theorem 1 *For the geometric order size distribution, the replenishment order process, N, is a reversible stochastic process whose stationary distribution is given by (2.3) and (2.5).*

Theorem 2 *If the order size distribution satisfies $p_k > 0$ for all $k = 1, 2, ...,$ then the replenishment order process, N, is a reversible stochastic process for all positive values of S if and only if the order size distribution is geometric.*

2.4 The Stationary Distribution of the Number of Units on Order

In this section we derive an explicit formula for the stationary distribution of the number of units on order in the lost sales model with stuttering Poisson demand. We consider only the partial fill case. A combinatorial argument is required (proof of Proposition 3) to collapse the state space. The complete fill case is treated in the appendix.

2.4.1 The Partial Fill Case

Let s index the number of units on order in the lost sales model, $s = 0, 1, ..., S$, Let $\pi = (\pi_s)$ denote the stationary distribution of the number of units on order.

We first derive an intermediate quantity. Let $\eta_{m,s}$ denote the stationary probability of the system having m orders outstanding and s units on order:

$$\eta_{m,s} = \sum_{\substack{i \in V \\ S - n_0(i) = s \\ m(i) = m}} \xi_i. \tag{2.6}$$

Letting $\bar{\nu} = \frac{1}{1 + \sum_{j \neq i_0} \nu_j}$, substitution from (2.3) and (2.5) yields

$$\begin{aligned} \eta_{m,s} &= \bar{\nu} \frac{(1-p)^s}{p^{1\{s=S\}}} \left(\frac{\lambda p}{\mu(1-p)}\right)^m \sum_{\substack{i \in V \\ S - n_0(i) = s \\ m(i) = m}} \frac{1}{\prod_{k=1}^{S} (n_k(i)!)} \qquad (2.7) \\ &= \bar{\nu} \frac{\left(\frac{\lambda}{\mu}\right)^m}{p^{1\{s=S\}}} p^m (1-p)^{s-m} \sum_{\substack{i \in V \\ S - n_0(i) = s \\ m(i) = m}} \frac{1}{\prod_{k=1}^{S} (n_k(i)!)}. \end{aligned}$$

Let $f_{NB}(\cdot; m, p)$ denote the negative binomial probability distribution with parameters m and p :

$$f_{NB}(x; m, p) \equiv \binom{m+x-1}{x} p^m (1-p)^x \text{ for} x = 0, 1, 2, \ldots.$$

Proposition 3 *For the lost sales model with stuttering Poisson demand and partial fills, the stationary probability of the system having m orders outstanding and s units on order is given by*

$$\eta_{m,s} = \bar{\nu} \frac{\left(\frac{\lambda}{\mu}\right)^m}{p^{1\{s=S\}}} \frac{f_{NB}(s-m; m, p)}{m!}.$$

Corollary 1 *For the lost sales model with stuttering Poisson demand and partial fills, the stationary distribution of the number of units on order is given by*

$$\pi_s = \frac{\sum_{m=0}^{s} \frac{\left(\frac{\lambda}{\mu}\right)^m}{p^{1\{s=S\}}} \frac{f_{NB}(s-m;m,p)}{m!}}{G(S)}, \tag{2.8}$$

where $G(S) = \sum_{s=0}^{S} \sum_{m=0}^{s} \frac{(\frac{\lambda}{\mu})^m}{m!} p^{-1_{\{s=S\}}} f_{\mathrm{NB}}(s-m; m, p)$, *and* $f_{\mathrm{NB}}(s-m; 0, p) = 1\{s = 0\}$ *when* $m = 0$.

Theorem 3 *For the lost sales model with stuttering Poisson demand, suppose the replenishment order lead times are independent and identically distributed and have a general distribution with finite mean* $\tau = \frac{1}{\mu}$, *with no point mass at zero. For the partial fill case, the stationary distribution of the number of units on order is given by*

$$\hat{\pi}_s = \pi_s,$$

where π_s *is the stationary distribution of the number of units on order in the lost sales model where lead times are independently identically exponential with mean* $\frac{1}{\mu}$ *respectively.*

The exact stationary distribution of the number of units-on-order (2.8) is the major contribution of this paper.

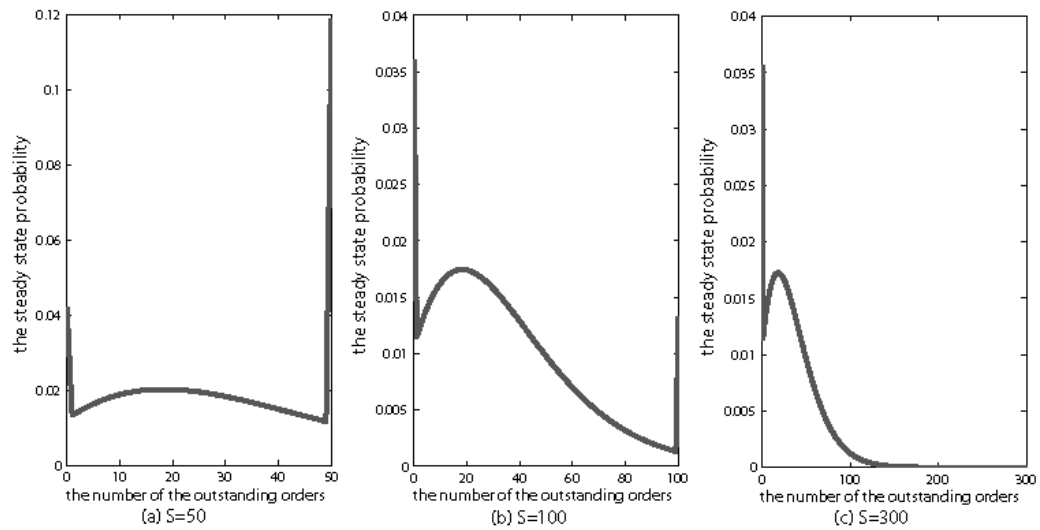

Figure 2.2: Distribution of Units on Order for the Stuttering Poisson Arrival Process

2.4.2 The Shape of the Units-on-Order Distribution in the Partial Fill Case

For a lost sales model with Poisson demand, the steady state distribution of the number of units on order is given by

$$\pi_s = \frac{\frac{e^{-\frac{\lambda}{\mu}}\left(\frac{\lambda}{\mu}\right)^s}{s!}}{\sum_{k=0}^{S}\frac{e^{-\frac{\lambda}{\mu}}\left(\frac{\lambda}{\mu}\right)^k}{k!}}, \tag{2.9}$$

(Muckstadt 2005 p44.). The basic unimodal shape does not change as a function of S.

Figure 2.2 (a),(b), and (c) are plots of the steady state distribution of the number of units on order in the stuttering Poisson case (partial fill) for different values of $S = 50, 100, 300$, respectively. The mean of the demand per unit time is 5 and the variance per unit time is 100. The distribution is trimodal with additional atoms occurring at 0 and S. The lead time mean is 7. Observe that, unlike the Poisson-based distribution, the atom at S becomes more pronounced as S decreases.

2.5 Comparison with F-S Results for the Partial Fill Case

Feeney and Sherbrooke (1966) discuss the compound Poisson demand process and give the stationary distribution for lost sales with partial fills allowed. With one exception (Baganha, 1985) this result has been unchallenged for forty years.

Let us restate their formula as follows by substituting for $\{y, x, s, T\}$ in the original paper with the notation $\{m, s, S, 1/\mu\}$ in the current paper. Then, using our notation, their formulas become

$$\begin{aligned} h(s) &= \frac{\sum_{m=0}^{s}((\frac{\lambda}{\mu})^m e^{-\frac{\lambda}{\mu}}/m!)f^{*m}(s)}{H(S)}, \text{for } 0 \leq s < S; \\ h(S) &= \frac{\sum_{m=0}^{S}((\frac{\lambda}{\mu})^m e^{-\frac{\lambda}{\mu}}/m!)\sum_{i=S}^{\infty} f^{*m}(i)}{H(S)}, \end{aligned} \tag{2.10}$$

where f^{*m} is the m-fold convolution of the order size distribution and $H(S)$ is the normalizer. These do not agree with (2.8) when $f(\cdot)$ is given by the geometric distribution. The difference can be traced to the reduced balance equations (A.7) in their paper(1966). Baganha(1985) noted that (A.7) is inconsistent with the proposed solution (A.8). However, even when corrected (A.7 in Baganha, 1985), these balance equations are built upon an implicit assumption in the F-S derivation that the distribution of order sizes in resupply is the same as the distribution of order sizes in customer arrivals. Since customer orders are filtered by the lost sales process, this assumption means that their analysis is not exact. Their result for the complete fill case does agree with (3.4), but does not actually solve their steady state equations. As we have shown, exact analysis is possible for the special case of stuttering Poisson demand. However, an exact analysis of the steady state distribution of units on order in the case of lost sales with general compound Poisson demand remains an open question.

In this section, we consider the quality of the F-S result as an approximation.

2.5.1 Analytical Comparison

The following theorem shows that the F-S formula for the stuttering Poisson demand process always overestimates the out-of-stock probability when the targeted inventory level S exceeds 1.

Theorem 4 *For $s = 0, 1, \ldots, S-1$,*

$$\pi_s G(S) e^{-\frac{\lambda}{\mu}} = h(s)H(S); \tag{2.11}$$

For $s = S = 1$, then

$$\pi_S G(S) e^{-\frac{\lambda}{\mu}} = h(S)H(S);$$

and if $s = S > 1$, then

$$\pi_S G(S) e^{-\frac{\lambda}{\mu}} < h(S)H(S). \tag{2.12}$$

2.5.2 Computational Comparison

We now consider the long run cost implications of using the approximate F-S model, (2.10), to optimize stock levels rather than the exact model (2.8). The exact model is used to evaluate the solutions.

Let c_h and c_p denote the holding cost per unit time and lost sales penalty per unit, respectively. Let $C(S)$ denote the long run average sum of holding costs

Table 2.1: Relative error of Feeney-Sherbrooke approximation when the ratio $\frac{\lambda}{\mu}$ changes:

$\frac{\lambda}{\mu}$	p	$\frac{c_p}{c_h}$	S_T^*	$C_T(S_T^*)$	S_A^*	$C_T(S_A^*)$	$\frac{C_T(S_A^*)-C_T(S_T^*)}{C_T(S_T^*)}$
0.7	0.4	10	1	2.5	1	2.5	0
1.4	0.4	10	3	4.3276	3	4.3276	0
2.1	0.4	10	5	5.772	6	5.8255	0.927%
2.8	0.4	10	7	7.0021	9	7.1418	1.99%
3.5	0.4	10	10	8.0842	12	8.3518	3.31 %
4.2	0.4	10	12	9.0512	14	9.2651	2.36 %
4.9	0.4	10	14	9.943	17	10.3204	3.79%
5.6	0.4	10	16	10.774	20	11.3528	5.37%
6.3	0.4	10	18	11.555	22	12.0715	4.46%
7	0.4	10	20	12.294	25	13.0411	6.07 %

and lost sales costs per unit time:

$$\begin{aligned} C(S) &= \textstyle\sum_{s=0}^{S} c_h(S-s)\pi_s + \sum_{s=0}^{S}[\sum_{j=1}^{\infty} c_p\lambda jp(1-p)^{(j-1)+(S-s)}]\pi_s \\ &= \textstyle\sum_{s=0}^{S}[c_h(S-s) + c_p\lambda\frac{(1-p)^{S-s}}{p}]\pi_s. \end{aligned}$$

Let $C_T(S)$ denote the time cost obtained using (2.8) and let $C_A(S)$ denote the approximate cost obtained using the F-S approximation (2.10). Let S_T^* denote the optimizer of $C_T(S)$ and S_A^* the optimizer of $C_A(S)$.

We use numerical methods to find S_T^* and S_A^*. We also compare $C_T(S_T^*)$ and $C_T(S_A^*)$, which are the true costs under optimized values. In Tables 2.1, 2.2, and 2.3, we fix the mean of the lead time $\tau = \frac{1}{\mu} = 7$ and vary λ, p and $\frac{c_p}{c_h}$ to study their effects on the difference between the costs obtained using the exact and the F-S models. We observe that the penalty cost of using the F-S model is small, less than 6% in most cases. Consequently it is unlikely that using the F-S

Table 2.2: Relative error of Feeney-Sherbrooke approximation when the probability p changes:

$\frac{\lambda}{\mu}$	p	$\frac{c_p}{c_h}$	S_T^*	$C_T(S_T^*)$	S_A^*	$C_T(S_A^*)$	$\frac{C_T(S_A^*)-C_T(S_T^*)}{C_T(S_T^*)}$
3.5	0.1	10	37	34.347	44	34.887	1.57 %
3.5	0.2	10	19	16.851	23	17.2607	2.43 %
3.5	0.3	10	13	11.011	15	11.2027	1.74%
3.5	0.4	10	10	8.0842	12	8.3518	3.31 %
3.5	0.5	10	8	6.3131	9	6.405	1.46 %
3.5	0.6	10	7	5.1412	8	5.2857	2.81 %
3.5	0.7	10	6	4.2819	6	4.2819	0
3.5	0.8	10	5	3.6273	5	3.6273	0
3.5	0.9	10	5	3.1445	5	3.1445	0
3.5	1	10	4	2.7123	4	2.7123	0

formula rather than the exact one is problematic. We further observe that when $\frac{\lambda}{\mu}$ increases, or VTMR= $\frac{2-p}{p}$ increases, the difference becomes more significant. However, when $\frac{c_p}{c_h}$ increases, the difference seems to fluctuate.

We conclude from this analysis that the F-S model provides a reasonably good approximation for the purpose of stock optimization, at least for the stuttering Poisson case.

2.6 Behavior of the Expected Ordered Unit Delivery Times

In this section we consider the deliveries of units on order in the lost sales model with partial fill and show that for the stuttering Poisson demand process, these

Table 2.3: Relative error of Feeney-Sherbrooke approximation when the ratio $\frac{c_p}{c_h}$ changes:

$\frac{\lambda}{\mu}$	p	$\frac{c_p}{c_h}$	S_T^*	$C_T(S_T^*)$	S_A^*	$C_T(S_A^*)$	$\frac{C_T(S_A^*)-C_T(S_T^*)}{C_T(S_T^*)}$
3.5	0.4	1	0	1	0	1	0
3.5	0.4	10	10	8.0842	12	8.3518	3.31 %
3.5	0.4	20	14	11.154	16	11.448	2.64 %
3.5	0.4	30	16	13.002	18	13.2427	1.85%
3.5	0.4	40	18	14.322	20	14.689	2.56%
3.5	0.4	50	19	15.33	21	15.656	2.12%
3.5	0.4	60	20	16.153	22	16.501	2.16%
3.5	0.4	70	21	16.851	22	16.987	0.8%
3.5	0.4	80	21	17.448	23	17.661	1.22%
3.5	0.4	90	22	17.957	24	18.306	1.95%

deliveries become more concentrated in time as the variance-to-mean ratio of the demand process increases.

This is not a surprising result, as we see in the following example. Consider a lost sales model with an $(S-1, S)$ policy where $S = 10$. Lead times are exponentially distributed with rate μ (the value of μ is irrelevant). Compare two demand processes that have identical expected rates of demand: in the first process, demand follows a Poisson process with rate $\lambda = 1$; in the second process, demand follows a compound Poisson process in which orders arrive at rate $\lambda = 0.01$ but each order is for exactly 100 units. Observe that cumulative expected demand over any constant length of time is the same in both cases. On the other hand, the variance of demand is higher when demand follows the compound Poisson process. If we observe the lost sales system at a random point in time, the probability distribution of the number of units on order of the Poisson system is given by (2.9). Furthermore, each unit on order corresponds to a unique customer order and each, therefore, belongs to a unique replenishment order. Consequently, conditioned on the number of units on order, s, the memoryless property of the exponential lead time distribution ensures that the deliveries of these s units will be spread out in time according to a distribution we will consider in detail in the sequel. For the extreme compound Poisson process, however, it is clear that any arriving customer demand order will always exceed the available stock. The number of units on order will be either 0 or 10 due to the partial fill assumption and the units on order will all belong to a single replenishment order. Thus, the units on order will always arrive together in a single delivery.

Intuitively, this is the limiting distribution of unit deliveries as the variance-to-mean ratio increases: all units arrive in a single delivery after an exponentially distributed lead time. Observe that, under both systems, the expected lead time, for an order, is $1/\mu$. It is the spread about this mean of individual unit deliveries that is of interest. From a managerial perspective, the two systems will behave very differently. In the Poisson system, if you are out of stock you can expect to receive a delivery of at least one unit in one-tenth of a lead time ($= 1/(S\mu)$). In the extreme version of the compound Poisson system, if you are out of stock you can expect to wait a full lead time ($= 1/\mu$) before seeing any units arrive. It is important for service parts planners to understand this phenomenon because, typically, variance-to-mean ratios are higher in the service parts industry than in consumer products environments. As one service parts manager expressed it, "the bad news is worse than I thought: if I am out of stock, I can expect to be out for a long time."

For the balance of this section, we focus on the stuttering Poisson demand process with parameters λ and p. Over any fixed length of time, T, the expected demand is $\lambda T/p$ and the variance of demand is $\lambda\left(\frac{2-p}{p^2}\right)T$. Denote the variance-to-mean ratio by $VTMR = \frac{2-p}{p}$. For a constant mean rate of demand, $\lambda/p = \bar{R}$, we investigate the impact of increasing the $VTMR$. That is, we consider the impact of letting $p \to 0$ while keeping $\lambda = \bar{R}p$.

We are interested in the spread of delivery times. Let O denote the number of units on order in steady state. Conditioned on $O = s$, let $M(s)$ denote the number of orders outstanding. Let t_m^s denote the remaining time until delivery of the m^{th} order, $m = 1, ..., M(s)$. Under the assumptions of the model, these remaining delivery times are independent, exponentially distributed random variables with mean $1/\mu$. Let their order statistics be denoted by $t_{(h)}^s$. In particular, $t_{(1)}^s$ is the remaining time until the delivery of the earliest order and $t_{(M(s))}^s$ is the remaining time until the delivery of the latest order. Let $\Delta(s) = \Delta_{\lambda,p,\mu,S}(s) \equiv E\left[t_{(M(s))}^s - t_{(1)}^s | O = s\right]$, the expected difference in steady state between the earliest and latest order delivery times, conditioned on s units outstanding. Then $\Delta(s)$ is a measure of the spread of order delivery times, in steady state, as a function of the parameters of the system. We show that $\Delta(s) \to 0$ monotonically as $p \to 0$ while keeping $\lambda = \bar{R}p$.

There are three steps to obtaining the result. The first is to show that $E\left[t_{(m)}^s - t_{(1)}^s | M(s) = m, O = s\right]$ is non-decreasing in m. The second is to show that $M(s)$ is stochastically decreasing as $p \to 0$ with $\lambda = \bar{R}p$. The third step is to show that $P\{M(s) > 1\}$ converges to 0 as $p \to 0$ with $\lambda = \bar{R}p$. First, we have the following result.

Lemma 1 *If $\{t_1, t_2, ..., t_m\}$ are independent and exponentially distributed, each with rate*

μ, then, for $h = 1, 2, ..., m$

$$E\left[t_{(h)}\right] = \sum_{k=m-h+1}^{m} \frac{1}{k\mu},$$

where $t_{(h)}$ is the h^{th} order statistic.

From this lemma, it follows that

$$E\left[t^s_{(m)} - t^s_{(1)} | M(s) = m, O = s\right] = \sum_{k=1}^{m} \frac{1}{k\mu} - \frac{1}{m\mu} = \sum_{k=1}^{m-1} \frac{1}{k\mu}$$

which is non-decreasing in m.

Let $\eta_{m,s}(p)$ denote the stationary probability with parameter p of having m orders outstanding and s units on order when p is the order size parameter. The distribution of $M(s)$ is given by

$$P\{M(s) = m\} = \eta_{m|s}(p) \equiv \frac{\eta_{m,s}(p)}{\sum_{h=1}^{s} \eta_{h,s}(p)}.$$

To show that this distribution is stochastically decreasing in p, we focus on the ratio of the successive probabilities

$$\begin{aligned} r^s_m(p) &\equiv \frac{\eta_{m,s}(p)}{\eta_{m-1,s}(p)} \\ &= \frac{\lambda}{\mu m} \frac{p}{1-p} \frac{s-m+1}{m-1}, \end{aligned}$$

which is decreasing as $p \to 0$ while keeping $\lambda = \bar{R}p$.

Lemma 2 *Suppose we have two random variables, M_1 and M_2, that take values in $\{1, 2, ..., m\}$ with probabilities $P\{M_h = l\} = f^{(h)}_l > 0$ for $h = 1, 2$ and $l = 1, 2, ..., m$. If*

$$\frac{f^{(1)}_{l+1}}{f^{(1)}_l} \geq \frac{f^{(2)}_{l+1}}{f^{(2)}_l},$$

then $P\{M_1 > l\} \geq P\{M_2 > l\}$ for all $l = 1, 2, ..., m$; that is, M_1 is stochastically greater than M_2.

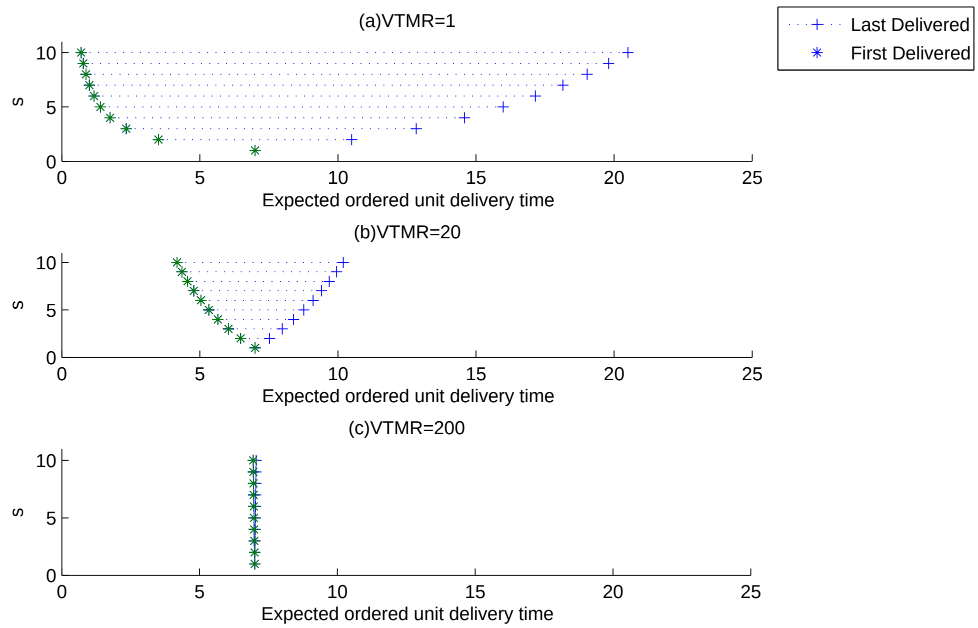

Figure 2.3: The Expected Spread of Deliveries

We now establish the following results.

Proposition 4 *For the lost sales model with stuttering Poisson demand and exponentially distributed lead times, $M(s)$, the number of outstanding orders in steady state, conditioned on s units on order, is stochastically decreasing as $p \to 0$ while keeping $\lambda = \bar{R}p$. Furthermore, $P\{M(s) > 1\}$ converges to 0 as $p \to 0$ with $\lambda = \bar{R}p$.*

The main result of this section is as follows.

Theorem 5 *For any given $s = 1, 2, ..., S$, the expected spread of deliveries in steady state approaches 0 monotonically as $p \to 0$, when $\lambda = \bar{R}p$.*

Corollary 2 *As $p \to 0$, the expected ordered unit delivery times, under the condition that s units are on-order, will converge to $E(t^1_{(1)}) = \frac{1}{\mu}$, while keeping $\lambda = \bar{R}p$.*

Figure 2.3 shows the convergence of the expected spread of deliveries as the VTMR goes from 1 to 200 for the case of $\mu = 5$, $L = 7$ and $S = 10$. The points on the left side of each graph are the expected times of the first delivered order with s units outstanding, $E(t^s_{(1)}|O = s)$, and those on the right side are the expected times of the last delivered order with s units outstanding, $E(E(t^s_{(m)}|O = s, M(s) = m)$.

2.7 Conclusion

In this paper, we conducted an exact analysis of the lost sales model with stuttering Poisson demand and exponentially distributed lead times under the $(S-1, S)$ inventory policy. We derived formulas to calculate the exact stationary distribution of the number of outstanding orders. This result was used to correct the long-standing more general result of Feeney and Sherbrooke for the stuttering Poisson case. We then demonstrated empirically that, at least for the stuttering Poisson case, the Feeney and Sherbrooke formulas are a good approximation (for partial fills) or exact (for complete fills) when used to set optimal stock levels. We also proved an interesting result that the spread of expected order replenishment delivery times becomes more concentrated as the $VTMR$ increases. The spread converges to zero around a single point, the mean of the lead time.

In a companion paper we use this lost sales model as the basis for modeling emergency order systems. We develop exact expressions for the first and second moments of the number of outstanding emergency orders and use them to estimate the mean, variance and atom at zero of the number of emergency units on order at the ESL.

2.A Proofs

Proof of Proposition 2:

For any path chosen according to the largest subscript rule and for the generators (2.1),

$$
\begin{aligned}
A_{i_0 j_{m(i)-1}} A_{j_{m(i)-1} j_{m(i)-2}} \ldots A_{j_1 i} &= \prod_{l=m(i),m(i)-1,\ldots,1} (\lambda p^{1\{n_0(j_{l-1})>0\}} (1-p)^{(k_{j_l,j_{l-1}}-1)}) \\
&= (\tfrac{\lambda p}{(1-p)})^{m(i)} (1-p)^{\sum_{l=1}^{m(i)} k_{j_l,j_{l-1}}} p^{-1\{n_0(i)=0\}} \\
&= (\tfrac{\lambda p}{(1-p)})^{m(i)} (1-p)^{S-n_0(i)} p^{-1\{n_0(i)=0\}}.
\end{aligned}
$$

Considering the path from i to i_0 and noting that if $n_k(i) = 0$, $n_k(i)! = 1$, we get

$$
A_{ij_1} A_{j_1 j_2} \ldots A_{j_{m(i)-1} i_0} = \prod_{l=0,1,\ldots,m(i)-1} \mu n_{k_{j_l,j_{l+1}}}(j_l) = \mu^{m(i)} \prod_{k=1}^{S} (n_k(i)!).
$$

Therefore, from (2.4)

$$
\nu_i \equiv \frac{A_{i_0 j_{m(i)-1}} A_{j_{m(i)-1} j_{m(i)-2}} \ldots A_{j_1 i}}{A_{ij_1} A_{j_1 j_2} \ldots A_{j_{m(i)-1} i_0}} = \frac{\left(\frac{\lambda p}{\mu(1-p)}\right)^{m(i)}}{\prod_{k=1}^{S} (n_k(i)!)} \frac{(1-p)^{S-n_0(i)}}{p^{1\{n_0(i)=0\}}}.
$$

■

Proof of Theorem 1:

Consider any two distinct states $i, i' \in V$, with $n(i) = (n_1(i), n_2(i), \ldots, n_S(i))$ and $n(i') = (n_1(i'), n_2(i'), \ldots, n_S(i'))$, $i \neq i'$.

1. Since $i \neq i'$, $\|(i, i')\| \neq 0$. Suppose $A_{ii'} = 0$, then by (2.1) $A_{i'i} = 0$, too. Hence, if $A_{ii'} = 0$, we have $\nu_i A_{ii'} = \nu_{i'} A_{i'i} \equiv 0$.

2. When $A_{ii'} \neq 0$ and $i \neq i'$, then, by (2.1), $\|(i,i')\| = 1$, and either $(i,i') \in V_C^2$ or $(i,i') \in V_R^2$. Without loss of generality, we assume $(i,i') \in V_C^2$ and $(i',i) \in V_R^2$. There are two subcases:

   - $n_0(i') > 0$: In this case, a demand of size $k_{ii'}$ arrives which is strictly less than $n_0(i)$. Then $A_{ii'} = \lambda p(1-p)^{(k_{ii'}-1)}$, $A_{i'i} = n_{k_{ii'}}(i')\mu = (n_{k_{ii'}}(i)+1)\mu$ and $m(i') = m(i)+1$. Hence

$$
\begin{aligned}
\nu_i A_{ii'} &= \frac{(\frac{\lambda p}{\mu(1-p)})^{m(i)}}{\prod_{k=1}^{S}(n_k(i)!)} \frac{(1-p)^{S-n_0(i)}}{p^{1\{n_0(i)=0\}}} \cdot (\lambda p(1-p)^{(k_{ii'}-1)}) \\
&= \frac{(\frac{1}{\mu(1-p)})^{m(i)}}{\prod_{k=1}^{S}(n_k(i)!)} \cdot (1-p)^{S-n_0(i)+k_{ii'}-1}(\lambda p)^{1+m(i)} \\
&= \frac{\mu(\frac{1}{\mu(1-p)})^{m(i')}}{\prod_{k=1}^{S}(n_k(i')!)} \frac{n_{k_{i,i'}}(i')!}{n_{k_{i,i'}}(i)!} \cdot (1-p)^{S-n_0(i')}(\lambda p)^{m(i')} \\
&= \frac{(\frac{\lambda p}{\mu(1-p)})^{m(i')}}{\prod_{k=1}^{S}(n_k(i')!)} \cdot (1-p)^{S-n_0(i')} \cdot (\mu\frac{n_{k_{i,i'}}(i')!}{n_{k_{i,i'}}(i)!}) \\
&= \nu_{i'} \cdot (n_{k_{i,i'}}(i')\mu) \\
&= \nu_{i'} A_{i'i}.
\end{aligned}
$$

   - $n_0(i') = 0$: In this case, a demand arrives and the demand size is equal to or greater than $k_{ii'} = n_0(i)$, so $A_{ii'} = \lambda(1-p)^{(k_{ii'}-1)}$, and $A_{i'i} = n_{k_{ii'}}(i')\mu$. But

$$
\nu_{i'} = \frac{\left(\frac{\lambda p}{\mu(1-p)}\right)^{m(i')}}{\prod_{k=1}^{S}(n_k(i')!)} \frac{(1-p)^{S-n_0(i')}}{p}.
$$

Similarly,

$$\begin{aligned}
\nu_i(A_{ii'}p) &= \frac{(\frac{\lambda p}{\mu(1-p)})^{m(i)}}{\prod_{k=1}^{S}(n_k(i)!)}\frac{(1-p)^{S-n_0(i)}}{p^{1\{n_0(i)=0\}}}\cdot(\lambda(1-p)^{(k_{ii'}-1)})p \\
&= \frac{(\frac{1}{\mu(1-p)})^{m(i)}}{\prod_{k=1}^{S}(n_k(i)!)}\cdot(1-p)^{S-n_0(i)+k_{ii'}-1}(\lambda p)^{1+m(i)} \\
&= \frac{\mu(\frac{1}{\mu(1-p)})^{m(i')}}{\prod_{k=1}^{S}(n_k(i')!)}\frac{n_{k_{i,i'}}(i')!}{n_{k_{i,i'}}(i)!}\cdot(1-p)^{S-n_0(i')}(\lambda p)^{m(i')} \\
&= \frac{(\frac{\lambda p}{\mu(1-p)})^{m(i')}}{\prod_{k=1}^{S}(n_k(i')!)}\cdot(1-p)^{S-n_0(i')}\cdot(\mu\frac{n_{k_{i,i'}}(i')!}{n_{k_{i,i'}}(i)!}) \\
&= \nu_{i'}p\cdot(n_{k_{i,i'}}(i')\mu) \\
&= (\nu_{i'}p)A_{i'i}.
\end{aligned}$$

Hence, $\nu_i A_{ii'} = \nu_{i'}A_{i'i}$.

Therefore, for any $i, i' \in V$, we have $\nu_i A_{ii'} = \nu_{i'}A_{i'i}$ and, after normalization, $\eta_i A_{ii'} = \eta_{i'}A_{i'i}$. By Proposition 1, N is a reversible stochastic process whose stationary distribution is given by (2.3) and (2.5). ∎

Proof of Theorem 2:

Let x and y be positive integers such that $S = x + y$ and $P(X = x) > 0$ and $P(X = y) > 0$. Consider the special states

$$n(i_0) = (0, 0, \ldots, 0)$$

and

$$\{n(i_2) : n_x(i_2) = 1, n_y(i_2) = 1, n_k(i_2) = 0, \text{for} k \neq x, y\}.$$

Now pick the cyclic sequence: $i_0 \to i_1 \to i_2 \to i'_1 \to i_0$, where

$$\{n(i_1) : n_x(i_1) = 1, n_k(i_1) = 0, \text{for} k \neq x\},$$

and

$$\{n(i'_1) : n_y(i'_1) = 1, n_k(i'_1) = 0, \text{for} k \neq y\}.$$

If this is a reversible Markov process,

$$\nu_{i_0} A_{i_0,i_1} A_{i_1,i_2} A_{i_2,i'_1} A_{i'_1,i_0} = \nu_{i_0} \lambda^2 p_x P(X \geq y) \mu^2$$

must equal

$$\nu_{i_0} A_{i_0,i'_1} A_{i'_1,i_2} A_{i_2,i_1} A_{i_1,i_0} = \nu_{i_0} \lambda^2 p_y P(X \geq x) \mu^2.$$

i.e.

$$p_x P(X \geq y) = p_y P(X \geq x).$$

Now, if $\{p_k > 0, \text{ for } k = 1, 2, \ldots\}$, an inductive proof easily establishes $p_k = p_1(1 - p_1)^{k-1}$ by letting $x \equiv 1$. So X must be geometrically distributed with parameter $p = p_1$. Combined with Theorem (3.3), this is a sufficient and necessary condition for reversibility. ■

Proof of Proposition 3:

First, let us show that

$$\sum_{\substack{i \in V \\ S - n_0(i) = s \\ m(i) = m}} \frac{m!}{\prod_{k=1}^{S} (n_k(i)!)} = \begin{pmatrix} s-1 \\ m-1 \end{pmatrix}. \tag{2.13}$$

To better understand the combinatorial expressions, we recast the language from orders and order sizes into boxes and balls. We are considering placing

s balls (i.e. units on order) into m boxes (i.e. orders). Suppose we have placed the s balls and have used exactly m boxes. Let $n_k \in \{0, 1, 2, ..., s\}$ denote the number of boxes that contain exactly k balls, $k = 1, 2, ..., s$. We refer to n_k as the box size count for box size (equivalently, for ball count) k. Of the $m!$ permutations of boxes, we are interested only in sequences that are unique with respect to the number of balls in each box. Thus, for example, if k_j is the number of balls in box j, $j = 1, ..., m$, the sequence $(k_1, k_2, k_3) = (0, 1, 1)$ corresponds to two equivalent permutations of the boxes since boxes numbered 2 and 3 can be reversed in sequence without changing the vector (k_1, k_2, k_3). For a given vector of box size counts, $n \equiv (n_1, n_2, ..., n_s)$, the number of permutations of boxes that are unique with respect to box size (i.e. ball count), is given by:

$$\frac{m!}{\prod_{\substack{k\in\{1,\ldots,s\} \\ n_k>0}} n_k!} = \frac{m!}{\prod_{k=1}^{s} n_k!},$$

where equality comes from the convention that $0! = 1$. From this, it follows that the number of ways of assigning s balls to exactly m boxes and sequencing the boxes so that the sequence is unique by ball count is given by

$$\sum_{\substack{n=(n_1,n_2,\ldots,n_s) \\ \sum_{k=1}^{s} kn_k=s \\ \sum_{k=1}^{s} n_k=m}} \frac{m!}{\prod_{k=1}^{s} n_k!} = \sum_{\substack{i\in V \\ S-n_0(i)=s \\ m(i)=m}} \frac{m!}{\prod_{k=1}^{S} (n_k(i)!)}.$$

This is the left hand side of (2.13). Now, we consider the same combinatorial problem from a different perspective. If we take any sequence of balls and place dividers between some of them, we could then assign the balls between dividers to boxes in sequence. The placement of dividers would uniquely define a sequence of ball counts per box. To ensure that exactly m boxes were used (with positive ball counts in each) we would have to place exactly $m - 1$ dividers into different positions between the s balls. (Placing two dividers between the same two balls would imply an empty box, which is not allowed.) Note that only $s - 1$

positions are available in this partitioning process; therefore, it follows that the number of ways to place these dividers is given by

$$\begin{pmatrix} s-1 \\ m-1 \end{pmatrix}.$$

From this we get (2.13).

Therefore,

$$\begin{aligned}
\eta_{m,s} &= \bar{\nu}\frac{\left(\frac{\lambda}{\mu}\right)^m}{p^{1\{s=S\}}m!}p^m\left(1-p\right)^{s-m}\begin{pmatrix} s-1 \\ m-1 \end{pmatrix} \\
&= \bar{\nu}\frac{(\frac{\lambda}{\mu})^m}{m!}p^{-1_{\{s=S\}}}\begin{pmatrix} s-1 \\ s-m \end{pmatrix}p^m(1-p)^{s-m} \\
&= \bar{\nu}\frac{(\frac{\lambda}{\mu})^m}{m!}p^{-1_{\{s=S\}}}f_{\text{NB}}(s-m;m,p).
\end{aligned}$$

■

Proof of Theorem 3:

Theorem 7 shows that the stationary distribution of ν_i is unchanged if the lead time has the same mean $\frac{1}{\mu}$ but has a general distribution where the lead times are independently identically distributed. Therefore the stationary distribution of the number of units on order is still the same as that when lead times are exponentially distributed. ■

Proof of Theorem 4:

Recall that f is the pdf of a geometric distribution with parameter p. Then

$$f^{*m}(s) = f_{\text{NB}}(s-m;m,p).$$

Thus, $\pi_s G(S) e^{-\frac{\lambda}{\mu}} = h(s)H(S)$ holds for $s = 0, 1, \ldots, S-1$.

When $s = S = 1$,

$$\pi_1 G(1) e^{-\frac{\lambda}{\mu}} = (\frac{\lambda}{\mu}) e^{-\frac{\lambda}{\mu}} \frac{f(1)}{p} = (\frac{\lambda}{\mu}) e^{-\frac{\lambda}{\mu}} = h(S)H(S).$$

Suppose $s = S > 1$ and $i > S$. When $m > 1$

$$\frac{f^{*m}(i+1)}{f^{*m}(i)} = \frac{\begin{pmatrix} (i+1)-1 \\ (i+1)-m \end{pmatrix} p^m (1-p)^{i+1-m}}{\begin{pmatrix} i-1 \\ i-m \end{pmatrix} p^m (1-p)^{i-m}} = \frac{i}{i+1-m}(1-p) > (1-p).$$

This means $f^{*m}(i+1) > (1-p) f^{*m}(i)$ and

$$f^{*m}(i) > (1-p)^{i-S} f^{*m}(S).$$

Therefore, when $S > 1$ and $m > 1$,

$$\sum_{i=S}^{\infty} f^{*m}(i) > \sum_{i=S}^{\infty} (1-p)^{i-S} f^{*m}(S) = f^{*m}(S) \sum_{i=0}^{\infty} (1-p)^i = \frac{f^{*m}(S)}{p}.$$

Since $f^{*0}(i) = 0 \text{for} i > 0$ and $f^{*1}(S)/p = f(S)/p = \sum_{i=S}^{\infty} f(i)$, we see that for $S > 1$

$$\pi_S G(S) e^{-\frac{\lambda}{\mu}} = \sum_{m=0}^{S} ((\frac{\lambda}{\mu})^m e^{-\frac{\lambda}{\mu}}/m!) \frac{f^{*m}(S)}{p} < \sum_{m=0}^{S} ((\frac{\lambda}{\mu})^m e^{-\frac{\lambda}{\mu}}/m!) \sum_{i=S}^{\infty} f^{*m}(i) = h(S)H(S).$$

After normalization, we have

$$h(S) > \pi_S \text{and } h(s) < \pi_s, \text{for} s = 0, 1, \ldots, S-1,$$

when $S > 1$. Furthermore,

$$\frac{h(s)}{\pi_s} = \frac{h(s')}{\pi_{s'}},$$

provided $s, s' < S$. ■

Proof of Lemma 1:

Due to the memoryless property of the exponential distribution random variables, we have that $t_{(k)} - t_{(k-1)} \sim Exp\left((m-k+1)\mu\right)$, for $k = 1, 2, ..., m$, and these differences are independent (Feller 1971 p19, Proposition9). $\blacksquare$

Proof of Lemma 2:

We need to show that $P(M_1 > k) \geq P(M_2 > k)$, for $k \in \{1, 2, \ldots, m-1\}$. Let $R_k = \frac{f_{k+1}^{(1)}}{f_k^{(1)}}, k = 1, 2, \ldots, m-1$. Then

$$f_{k+1}^{(i)} = f_1^{(i)} R_1^{(i)} R_2^{(i)} \cdots R_k^{(i)},$$

and

$$1 = f_1^{(1)}(1 + R_1^{(1)} + \cdots + R_1^{(1)} R_2^{(1)} \cdots R_{m-1}^{(1)}) = f_1^{(2)}(1 + R_1^{(2)} + \cdots + R_1^{(2)} R_2^{(2)} \cdots R_{m-1}^{(2)}).$$

This implies that

$$\frac{f_1^{(2)}}{f_1^{(1)}} = \frac{1 + R_1^{(1)} + \cdots + R_1^{(1)} R_2^{(1)} \cdots R_{m-1}^{(1)}}{1 + R_1^{(2)} + \cdots + R_1^{(2)} R_2^{(2)} \cdots R_{m-1}^{(2)}} \geq 1,$$

since $R_k^{(1)} \geq R_k^{(2)}$.

For any value $k \in \{2, \ldots, m-1\}$, we have
$$\frac{R_1^{(1)} R_2^{(1)} \cdots R_{k-1}^{(1)} + \cdots + R_1^{(1)} R_2^{(1)} \cdots R_{m-1}^{(1)}}{R_1^{(2)} R_2^{(2)} \cdots R_{k-1}^{(2)} + \cdots + R_1^{(2)} R_2^{(2)} \cdots R_{m-1}^{(2)}} \leq \frac{R_1^{(1)} R_2^{(1)} \cdots R_k^{(1)} + \cdots + R_1^{(1)} R_2^{(1)} \cdots R_{m-1}^{(1)}}{R_1^{(2)} R_2^{(2)} \cdots R_k^{(2)} + \cdots + R_1^{(2)} R_2^{(2)} \cdots R_{m-1}^{(2)}}.$$
To obtain this ,we need to show that

$$R_1^{(1)} R_2^{(1)} \cdots R_{k-1}^{(1)} (R_1^{(2)} R_2^{(2)} \cdots R_k^{(2)} + \cdots + R_1^{(2)} R_2^{(2)} \cdots R_{m-1}^{(2)}),$$

is less than or equal to

$$R_1^{(2)} R_2^{(2)} \cdots R_{k-1}^{(2)} (R_1^{(1)} R_2^{(1)} \cdots R_k^{(1)} + \cdots + R_1^{(1)} R_2^{(1)} \cdots R_{m-1}^{(1)}).$$

But this follows immediately since $R_j^{(1)} \geq R_j^{(2)}$ for any j. Therefore,

$$\frac{1 + R_1^{(1)} + \cdots + R_1^{(1)} R_2^{(1)} \cdots R_{m-1}^{(1)}}{1 + R_1^{(2)} + \cdots + R_1^{(2)} R_2^{(2)} \cdots R_{m-1}^{(2)}} \leq \frac{R_1^{(1)} + \cdots + R_1^{(1)} R_2^{(1)} \cdots R_{m-1}^{(1)}}{R_1^{(2)} + \cdots + R_1^{(2)} R_2^{(2)} \cdots R_{m-1}^{(2)}} \leq \cdots \leq \frac{R_1^{(1)} R_2^{(1)} \cdots R_{m-1}^{(2)}}{R_1^{(2)} R_2^{(2)} \cdots R_{m-1}^{(2)}}.$$

By multiplying $\frac{f_1^{(1)}}{f_1^{(2)}}$ we have

$$1 = \frac{f_1^{(1)}}{f_1^{(2)}} \frac{1 + R_1^{(1)} + \cdots + R_1^{(1)} R_2^{(1)} \cdots R_{m-1}^{(1)}}{1 + R_1^{(2)} + \cdots + R_1^{(2)} R_2^{(2)} \cdots R_{m-1}^{(2)}} \leq \cdots \leq \frac{f_1^{(1)}}{f_1^{(2)}} \frac{R_1^{(1)} R_2^{(1)} \cdots R_{m-1}^{(2)}}{R_1^{(2)} R_2^{(2)} \cdots R_{m-1}^{(2)}}.$$

Restated,

$$1 = \frac{P(M_1 > 0)}{P(M_2 > 0)} \leq \frac{P(M_1 > 1)}{P(M_2 > 1)} \leq \cdots \leq \frac{P(M_1 > m-1)}{P(M_2 > m-1)}.$$

Hence, for any real value k,

$$P(M_1 > k) \leq P(M_2 > k).$$

$\blacksquare$

Proof of Proposition 4:

From lemma 2 we see that $M(s)$ is stochastically decreasing as $p \to 0$ when $\lambda = \bar{R}p$ and that $r_m^s(p)$ is decreasing in p for fixed λ/p.

For each s fixed, $r_m^s(p) = O(p)$ (that is, $\lim_{p\to 0} \frac{r_m^s(p)}{p}$ is a constant). Then we can write

$$\eta_{m,s}(p) = O(p)\eta_{1,s}(p), \text{for} m = 2, 3, \ldots, s.$$

Therefore, the conditional distribution

$$\eta_{m|s}(p) = \frac{\eta_{m,s}}{\sum_{m=1}^{s} \eta_{m,s}} = \frac{O(p)}{1 + (s-1)O(p)}, \text{for} m = 2, 3, \ldots, s,$$

and $\eta_{1|s}(p) \to 1$, as $p \to 0$. That is, $P\{M(s) > 1\}$ converges to 0 as $p \to 0$ with $\lambda = \bar{R}p$. $\blacksquare$

Proof of Theorem 4:

Since

$$\begin{aligned}\Delta(s) &= E(E\left[t^s_{(m)} - t^s_{(1)}|M(s) = m, O = s\right]) \\ &= \textstyle\sum_{m=1}^{s} \eta_{m|s}(p) E\left[t^s_{(m)} - t^s_{(1)}|M(s)m, O = s\right] \\ &= \textstyle\sum_{m=2}^{s} \frac{1}{\mu} \sum_{k=1}^{m-1} \frac{1}{k} P\{M(s) > 1\}.\end{aligned}$$

Combined with Proposition 4 and the stochastic ordering result if $A \underset{\sim}{\prec} B$ and u is a non-decreasing function then $E[u(A)] \leq E[u(B)]$ (Puterman 2005), $\Delta(s)$ is decreasing as p converges to 0 while keeping λ/p constant. ∎

Proof of Corollary 2:

By Proposition 4 and Theorem 5, the expected ordered unit delivery times, under the condition that s units are on-order, should be the same and equal to

$$E(t^s_{(1)}|O = s) = E(E(t^m_{(1)}|M(s) = m, O = s)) = \frac{1}{\mu}\eta_{1|s}(p) + \sum_{m=2}^{s} E(t^m_{(m)} - t^m_{(1)})\eta_{m|s}(p) \to \frac{1}{\mu}$$

as $p \to 0$. ∎

2.B The Complete Fill Case

To this point we have considered only the partial fill case. Another possibility is that a customer order is rejected (all units lost) if there is insufficient stock on hand to fill it completely. We refer to this as the complete fill case. The analysis is very similar to the partial fill case. We have

$$\eta_{m,s} = \eta_{0,0} \left(\frac{\lambda}{\mu}\right)^m \frac{f_{NB}(s - m; m, p)}{m!}, \tag{2.14}$$

where $\eta_{0,0}$ is the normalizer. The steady state distribution of units on order is given by the following:

Proposition 5 *For the lost sales model with stuttering Poisson demand and complete fills, the stationary distribution of the number of units on order is given by:*

$$\pi_s = \frac{\sum_{m=0}^{s} (\frac{\lambda}{\mu})^m \frac{f_{NB}(s-m;m,p)}{m!}}{G(S)}, \tag{2.15}$$

where $G(S) = \sum_{s=0}^{S} \sum_{m=0}^{s} \frac{(\frac{\lambda}{\mu})^m}{m!} f_{\mathrm{NB}}(s-m;m,p)$, *and* $f_{\mathrm{NB}}(s-m;0,p) = 1\{s=0\}$ *when* $m=0$. *i.e. the truncated compound Poisson distribution.*

Proof: In the case of complete fill the accepted demand is given by $X1_{X \le I}$, where X is the customer order size and I is inventory on hand at the time of the order, as before. The infinitesimal generator in the stuttering Poisson case becomes:

$$A_{ij} \equiv \begin{cases} n_{k_{ij}}(i)\mu & \text{if} (i,j) \in V_R^2, \\ \lambda p_{k_{ij}} & \text{if} (i,j) \in V_C^2, \\ -(m(i)\mu + \lambda(1 - \overline{P}_{n_0(i)}))1\{n_0(i) \neq 0\} & \text{if} \, j = i, \\ 0 & \text{otherwise.} \end{cases} \tag{2.16}$$

Following the notation and method of section 3, we get

$$v_i \equiv \frac{\left(\frac{\lambda p}{\mu(1-p)}\right)^{m(i)}}{\prod_{k=1}^{S} (n_k(i)!)} (1-p)^{S-n_0(i)} \tag{2.17}$$

as the complete fill counterpart to (2.5). Observe that the term $\frac{1}{p^{1\{n_0(i)=0\}}}$ is needed for the partial fill case (2.5).

In the analog of Theorem 1 for the complete fill case, simply replace (2.5) with (3.2). The proof is identical except that the case $n_0(i) = 0$ is no different from the $n_0(i) > 0$ case with complete fills. In the analog of Proposition 2 and Corollary 1, omit the factor $\frac{1}{p^{1\{n_0(i)=0\}}}$ or $\frac{1}{p^{1\{s=S\}}}$. The analog to (2.8) for the complete fill case becomes (3.4). ∎

This is a bimodal distribution because the mode at $s = S$ disappears.

Theorem 6 *Suppose in the lost sales model that demand occurs according to stuttering Poisson process and the replenishment order lead times are independent and identically distributed and have general distribution with finite mean $L = \frac{1}{\mu}$, where there is no point mass at zero. For the complete fill case, the stationary distribution of the number of units on order is given by*

$$\hat{\pi}_s = \pi_s,$$

where π_s, given by (3.4), is the stationary distribution of the number of units on order in the lost sales model when lead times are exponentially distributed with mean $\frac{1}{\mu}$.

The proof is similar to that of Theorem 3.

2.C A Lost Sales Model with Stuttering Poisson Demand and General Lead Times

We always assume that the lead times are independently identically distributed. The original process $X(t)$, with exponentially distributed lead times is a Markov process. When it is extended to the case of general time distributions, it becomes a generalized semi-Markov process (GSMP). By extending the state space, we can obtain a Markov process and derive the stationary distribution of the extended state space. Finally, we could prove the marginal stationary distribution of the number of units on order does not depend on the lead time distribution but only on its mean.

Let $F(\cdot)$ denote the general cumulative distribution function(CDF) of order lead times with no point mass at zero. Expand the underlying state space from V to $V \times \Re^{+}_{S,S}$. Here $U = (u_{s,r}) \in \Re^{+}_{S,S}$ is an S by S matrix with non-negative elements. We construct the lost sales model with generally distributed lead times as a stochastic process $Z(t)$, with state space $V \times \Re^{+}_{S,S}$:

$$Z(t) = (i, U) = (n_1(i), n_2(i), \ldots, n_S(i); U) = \begin{pmatrix} n_1(i) & n_2(i) & \ldots & \ldots & n_S(i) \\ u_1^{(1)} & u_2^{(1)} & \ldots & \ldots & u_S^{(1)} \\ u_1^{(2)} & u_2^{(2)} & \ldots & \ldots & u_S^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ u_1^{(S)} & u_2^{(S)} & \ldots & \ldots & u_S^{(S)} \end{pmatrix}.$$

Here, $n_s(i)$ is the number of outstanding orders with size s and $u_s^{(1)} \geq u_s^{(2)} \geq \ldots \geq u_s^{(S)} \geq 0$ stand for the ordered replenishment ages for orders with size s. That is $u_s^{(r)}$ is the age of the rth oldest replenishment order of size s. The new process $Z(t)$ is a Markov process on $V \times \Re^{+}_{S,S}$.

Define $R(i) = \{(s, r) : r \leq n_s(i)\}$ as the replenishment order index set. So we have $u_s^{(r)} = 0$ if $(s, r) \notin R(i)$. Define

$$\Re_{S,S}(i) \equiv \{U \in \Re^{+}_{S,S} : u_s^{(1)} \geq u_s^{(2)} \geq \ldots \geq u_s^{(S)} \geq 0, \text{ and } u_s^{(r)} \equiv 0 \text{ if } (s, r) \notin R(i)\}.$$

Each state (i, U) in this system satisfies the condition $U \in \Re_{S,S}(i)$ and therefore we have $(i, U) \in V \times \Re_{S,S}(i) \subseteq V \times \Re^{+}_{S,S}$.

Our intent is to show that the stationary distribution of $Z(t)$ is insensitive to the lead time distribution for a given mean, $\frac{1}{\mu}$, under the partial fill case. The proof for the complete fill case is nearly the same.

Lemma 3 *Given state i* $((n_1(i), n_2(i), \ldots\ldots n_S(i)))$*,*

$$\int_{U \in \Re_{S,S}(i)} \prod_{s,r} [1 - F(u_s^{(r)})] du_1^{(1)} \ldots u_S^{(S)} = \frac{1}{\prod_{s=1}^{S} n_s(i)!} (\frac{1}{\mu})^{m(i)},$$

where $m(i) = \sum_{s=1}^{S} n_s(i)$.

Proof: Since $\prod_{s,r}[1 - F(u_s^{(r)})]$ does not depend on the order of $u_s^{(r)}$, we could integrate it on the whole space and divide the results by $n_s(i)!$ for each s fixed. Therefore,

$$\begin{aligned}
& \int_{U \in \Re_{S,S}(i)} \prod_{s,r} [1 - F(u_s^{(r)})] du_1^{(1)} \ldots u_S^{(S)} \\
= \;& \prod_{s=1}^{S} [\int_{u_s^{(1)} \geq u_s^{(2)} \geq \ldots \geq u_s^{(S)} \geq 0} \prod_r [1 - F(u_s^{(r)})] du_s^{(1)} \ldots u_s^{(S)}] \\
= \;& \prod_{s=1}^{S} [\int_{t_{s,1}=0}^{\infty} \ldots \int_{t_{s,S}=0}^{\infty} [1 - F(t_{s,r})] \tfrac{1}{n_s(i)!} dt_{s,1} \ldots dt_{s,S}] \\
= \;& \prod_{s=1}^{S} \{\tfrac{1}{n_s(i)!} [\int_{t=0}^{\infty} [1 - F(t)] dt]^{n_s(i)}\} \\
= \;& \prod_{s=1}^{S} \{\tfrac{1}{n_s(i)!} [\tfrac{1}{\mu}]^{n_s(i)}\} \\
= \;& \tfrac{1}{\prod_{s=1}^{S} n_s(i)!} (\tfrac{1}{\mu})^{m(i)}.
\end{aligned}$$

■

The proof of uniqueness and ergodicity of the stationary distribution of this Markov process $Z(t)$ is a consequence of Theorem 1 in Sevastyanov (1957). The proof just follows the routine of proving the results for a telephone system with refusals (Sevastyanov, 1957, section 3). The stationary distribution of $Z(t)$ is given by the following theorem. The marginal distribution of $X(t)$ is seen to be invariant to the form of the lead time distribution.

Theorem 7 *The steady state distribution of this Markov process* $Z(t)$*,* $\zeta_{(i,U)}$ *is*

$$\zeta_{(i,U)} = C \left(\frac{\lambda p}{1 - p} \right)^{m(i)} \frac{(1 - p)^{S - n_0(i)}}{p^{1\{n_0(i)=0\}}} \prod_{(s,r) \in R(i)} [1 - F(u_s^{(r)})], \tag{2.18}$$

where $C = \frac{1}{G(S)}$ is the same normalizer as in Corollary 1. Therefore, the steady state distribution of the original GSMP,

$$\tilde{\xi}_i = \int_{U \in \mathfrak{R}^+_{S,S}} \zeta_{(i,U)} dU = C \frac{\left(\frac{\lambda p}{\mu(1-p)}\right)^{m(i)}}{\prod_{r=1}^{S}(n_r(i)!)} \frac{(1-p)^{S-n_0(i)}}{p^{1\{n_0(i)=0\}}},$$

which is the same stationary distribution that is obtained when the lead times are exponentially distributed with mean $\frac{1}{\mu}$.

Proof: Notice that $\prod_{(s,r)}[1-F(u_s^{(r)})] = \prod_{(s,r)\in R(i)}[1-F(u_s^{(r)})]$ since $1-F(u_s^{(r)}) = 1$ for (s,r) outside of $R(i)$. Integrating $\zeta_{(i,U)}$ with respect to $U \in \mathfrak{R}_{S,S}(i)$ and use Lemma3, we have

$$\begin{aligned}
\int_{U \in \mathfrak{R}_{S,S}(i)} \zeta_{(i,U)} dU &= C\left(\frac{\lambda p}{1-p}\right)^{m(i)} \frac{(1-p)^{S-n_0(i)}}{p^{1\{n_0(i)=0\}}} \int_{U \in \mathfrak{R}_{S,S}(i)} \prod_{s,r}[1-F(u_s^{(r)})]dU \\
&= C \frac{\left(\frac{\lambda p}{\mu(1-p)}\right)^{m(i)}}{\prod_{r=1}^{S}(n_r(i)!)} \frac{(1-p)^{S-n_0(i)}}{p^{1\{n_0(i)=0\}}}
\end{aligned}$$

Let $U+\Delta t$(or $U-\Delta t$) denote adding (or subtracting) small Δt (or $\min(\Delta t, u_s^{(r)})$) to U's each entry $u_s^{(r)}$ if $(s,r) \in R(i)$.

We claim that for Δt sufficiently small, there will occur at most one event (customer arrival or order replenishment delivery) in the interval $(t, t+\Delta t]$ for any t. This follows because the delivery process is simply a shifted, filtered version of the arrival process. Consequently, the combined process is a filtered version of a Poisson process (refer to Resnick 2005 , section 4.4 page 316). So now we choose a Δt sufficiently small so that at most one event happens within the interval $(t, t+\Delta t]$.

Define $Q_{(i,U),(j,U')}(\Delta t)$ as the transition probability from state (i,U) to state (j,U') during time Δt. Since $Z(t)$ is a Markov process, Q has no dependence on t. For sufficiently small Δt, we have the following transition probabilities:

Case 1 : If no customer arrives, $n_0(i) = S$, and $(j, U') = (i_0, O)$, where O is the matrix with zeros entries,

$$Q_{(i_0,O),(i_0,O)}(\Delta t) = 1 - \lambda\Delta t + o(\Delta t).$$

Case 2 : If no replenishment order arrives when (i, U) has $n_0(i) = 0$ (any arrival is lost), we have

$$Q_{(i,U),(i,U+\Delta t)}(\Delta t) = \prod_{(s,r)\in R(i)} \frac{1 - F(u_s^{(r)} + \Delta t)}{1 - F(u_s^{(r)})}.$$

Case 3 : If no customer arrives for general state (i, U) with $0 < n_0(i) < S$,

- (Case 3a) When no customer arrives, or no replenishment order arrives during time Δt case,

$$Q_{(i,U),(i,U+\Delta t)}(\Delta t) = \prod_{(s,r)\in R(i)} \frac{1 - F(u_s^{(r)} + \Delta t)}{1 - F(u_s^{(r)})}(1 - \lambda\Delta t + o(\Delta t))$$

- (Case 3b) Now suppose no customer arrives but one replenishment order of size k_{ij} ($(i, j) \in V_R^2$) arrives. Suppose that order is the lth oldest order, $1 \le l \le n_{k_{ij}}(i)$. Let $U_{i,j}^{l-}$ be the same as U except that the element $u_{k_{ij}}^{(l)}$ is deleted so that the lth column changes from

$$(u_{k_{ij}}^{(1)}, \ldots, u_{k_{ij}}^{(n_{k_{ij}}(i))}, 0, \ldots, 0)',$$

to

$$(u_{k_{ij}}^{(1)}, \ldots, u_{k_{ij}}^{(l-1)}, u_{k_{ij}}^{(l+1)}, \ldots, u_{k_{ij}}^{(n_{k_{ij}}(i))}, 0, 0, \ldots, 0)'.$$

Actually, U_{ij}^{l-} is U after recording delivery of lth oldest order of size k_{ij}. Thus,

$$\begin{aligned} & Q_{(i,U),(j,U_{i,j}^{l-}+\Delta t)}(\Delta t) \\ = \; & \tfrac{F(u_{k_{ij}}^{(l)}+\Delta t) - F(u_{k_{ij}}^{(l)})}{1-F(u_{k_{ij}}^{(l)}))} \prod_{(s,r)\in R(i)/(k_{ij},l)} \tfrac{1-F(u_s^{(r)}+\Delta t)}{1-F(u_s^{(r)})}(1 - \lambda\Delta t + o(\Delta t)). \end{aligned}$$

Case 4 : When (i, U) satisfies $n_0(i) > 0$, and one customer arrives with accepted order size k_{ij} ($(i, j) \in V_C^2$) and has age u ($0 < u \le \Delta t$) at the end of the interval and no replenishment order arrives, we have

$$Q_{(i,U),(j,U_{i,j})}(\Delta t) = \prod_{(s,r)\in R(i)} \frac{1 - F(u_s^{(r)} + \Delta t)}{1 - F(u_s^{(r)})}(A_{ij}\Delta t + o(\Delta t))(\frac{1}{\Delta t})(1 - F(u)).$$

Here $U_{i,j} = U + \Delta t$ except that the new replenishment order caused by the new arrival has age $u_{k_{ij}}^{(n_{k_{ij}}(j))} = u$. Notice that $(\frac{1}{\Delta t})$ is the conditional density of the new replenishment order with u being the age at the end of the interval $(0, \Delta t]$. This is because of the uniformly distributed arrival time of the Poisson process conditioned on one arrival occurring during an interval of length Δt. A special case when $(i, U) = (i_0, O)$,$(i_0, j) \in V_C^2$, we have

$$Q_{(i_0,O),(j,U_{i_0,j})}(\Delta t) = \frac{A_{i_0,j}\Delta t + o(\Delta t)}{\Delta t}(1 - F(u)),$$

where $U_{i_0,j} = O$ except $u_{k_{ij}}^{(1)} = u$.

Define $P_{(i,U)}(t) = P[Z(t) = (i, U)]$. Making use of the Markov property and $Q_{(i,U)(j,U')}(\Delta t)$, we obtain:

Case 1 For $(i, U) = (i_0, O)$,

$$\begin{aligned} & P_{(i_0,O)}(t + \Delta t) \\ = \quad & P_{(i_0,O)}(t)(1 - \lambda\Delta t) + \textstyle\sum_{\{j:(j,i_0)\in V_R^2\}} \int_0^\infty P_{(j,U)}(t) \frac{F(u_{k_{ji_0}}^{(1)} + \Delta t) - F(u_{k_{ji}}^{(1)})}{1 - F(u_{k_{ji_0}}^{(1)})} du_{k_{ji_0}}^{(1)} + o(\Delta t) \end{aligned} \tag{2.19}$$

except $u_{k_{ji_0}}^{(1)}$, the other entries of U are zeros.

Case 2 For (i, U) with $n_0(i) = 0$,

$$P_{(i,U)}(t + \Delta t) = P_{(i,U-\Delta t)}(t) \prod_{(s,r)\in R(i)} \frac{1 - F(u_s^{(r)})}{1 - F(u_s^{(r)} - \Delta t)} + o(\Delta t). \tag{2.20}$$

Case 3 For general (i, U) with $0 < n_0(i) < S$, and $u_s^{(r)} > 0$ for all $(s, r) \in R(i)$,

$$\begin{aligned}
& P_{(i,U)}(t + \Delta t) \\
= \;& P_{(i,U-\Delta t)}(t) \prod_{(s,r)\in R(i)} \frac{1-F(u_s^{(r)})}{1-F(u_s^{(r)}-\Delta t)}(1 - \lambda\Delta t) \\
& + \sum_{\{(j,U'):(j,i)\in V_R^2, U'=(U-\Delta t)_{ji}^+\}} \int_0^\infty P_{(j,U')}(t) \prod_{(s,r)\in R(i)} \frac{1-F(u_s^{(r)})}{1-F(u_s^{(r)}-\Delta t)} \frac{F(u)-F(u-\Delta t)}{1-F(u-\Delta t)} du(1 - \lambda\Delta t) \\
& + o(\Delta t),
\end{aligned} \tag{2.21}$$

where $(U - \Delta t)_{ji}^+$ is the $U - \Delta t$ inserting $u_{k_{ji}}^{(l)} = u$ for some $l \leq n_{k_{ji}}(j)$.

Case 4 For general (i, U), with one $u_s^{(r)} = u$ with $0 < u \leq \Delta t$ for $(s, r) \in R(i)$,

$$P_{(i,U)}(t + \Delta t) = P_{(j,U-\Delta t)}(t) \prod_{(s,r)\in R(j)} \frac{1 - F(u_s^{(r)})}{1 - F(u_s^{(r)} - \Delta t)}(A_{ji}\Delta t + o(\Delta t))\frac{1}{\Delta t}(1 - F(u)), \tag{2.22}$$

where $(j, i) \in V_C^2$.

Define $P_{(i,U)}^*(t) = \frac{P_{(i,U)}(t)}{\prod_{(s,r)\in R(i)}[1-F(u_s^{(r)})]}$, which is the conditional probability in state i given the ages of replenishment orders at time t. Assume the existence of $\frac{\partial P_{(i,U)}^*(t)}{\partial t}$ and $\frac{\partial P_{(i,U)}^*(t)}{\partial u_s^{(r)}}$. Dividing equations (2.19)-(2.21) by Δt and letting $\Delta t \to 0$ in equation (2.19)-(2.22), we obtain the following system of integro-differential equations

Case 1

$$\frac{\partial P_{(i_0,O)}^*(t)}{\partial t} + \lambda P_{(i_0,O)}^*(t) = \sum_{\{j:(j,i_0)\in V_R^2\}} \int_0^\infty P_{(j,U)}^*(t) dF(u_{k_{ji_0}}^{(1)}),$$

$u_{k_{ji_0}}^{(1)}$ is the only positive entry of U.

Case 2 For (i, U) with $n_0(i) = 0$,

$$\frac{\partial P_{(i,U)}^*(t)}{\partial t} + \sum_{(s,r)\in R(i)} \frac{\partial P_{(i,U)}^*(t)}{\partial u_s^{(r)}} = 0.$$

Case 3 For general (i, U) with $i \neq i_0$ and $0 < n_0(i) < S$,

$$\frac{\partial P_{(i,U)}^*(t)}{\partial t} + \sum_{(s,r)\in R(i)} \frac{\partial P_{(i,U)}^*(t)}{\partial u_s^{(r)}} + \lambda P_{(i,U)}^*(t) = \sum_{\{(j,U'):(j,i)\in V_R^2, U'=(U)_{ji}^+\}} \int_0^\infty P_{(j,U')}^*(t) dF(u) \tag{2.23}$$

where $(U)^+_{ji}$ is the U inserting $u^{(l)}_{k_{ji}} = u$ for some $l \le n_{k_{ji}}(j)$.

Case 4 for $(j, i) \in V^2_C$,

$$P^*_{(i,U)}(t) = A_{ji} P^*_{(j,U)}(t).$$

If we start with the stationary distribution, then all the derivatives with respect to time t vanish. Dropping the dependence on t, we have

Case 1

$$\lambda P^*_{(i_0,O)} = \sum_{\{j:(j,i_0)\in V^2_R\}} \int_0^\infty P^*_{(j,U)} dF(u^{(1)}_{k_{ji_0}}), \tag{2.24}$$

$u^{(1)}_{k_{ji_0}}$ is the only positive entry of U.

Case 2 For (i, U) with $n_0(i) = 0$,

$$\sum_{(s,r)\in R(i)} \frac{\partial P^*_{(i,U)}}{\partial u^{(r)}_s} = 0. \tag{2.25}$$

Case 3 For general (i, U) with $i \ne i_0$ and $0 < n_0(i) < S$,

$$\sum_{(s,r)\in R(i)} \frac{\partial P^*_{(i,U)}}{\partial u^{(r)}_s} + \lambda P^*_{(i,U)} = \sum_{\{(j,U'):(j,i)\in V^2_R, U'=(U)^+_{ji}\}} \int_0^\infty P^*_{(j,U')} dF(u) \tag{2.26}$$

where $(U)^+_{ji}$ is the U inserting $u^{(l)}_{k_{ji}} = u$ for some $l \le n_{k_{ji}}(j)$.

Case 4 For $(j, i) \in V^2_C$,

$$P^*_{(i,U)}(t) = A_{ji} P^*_{(j,U)}(t). \tag{2.27}$$

Let $\zeta_{(i,U)}$ be given by (2.18). It is straightforward to verify that the substitution $P^*_{(i,U)}$ by $\frac{\zeta_{(i,U)}}{\prod_{(s,r)\in R(i)}[1-F(u^{(r)}_s)]}$ satisfies equations (2.24)-(2.27). We show only the proof of Case 3 here. From (2.18), we know

$$\frac{\zeta_{(i,U)}}{\prod_{(s,r)\in R(i)}[1 - F(u^{(r)}_s)]} = C\left(\frac{\lambda p}{1-p}\right)^{m(i)} \frac{(1-p)^{S-n_0(i)}}{p^{1\{n_0(i)=0\}}}, \tag{2.28}$$

where C is the normalizer. Now the substitution $P^*_{(i,U)}$ in (2.26) by (2.28), we have the left hand side (LHS) as

$$LHS = \lambda C \left(\frac{\lambda p}{1-p}\right)^{m(i)} \frac{(1-p)^{S-n_0(i)}}{p^{1\{n_0(i)=0\}}} = \lambda C \left(\frac{\lambda p}{1-p}\right)^{m(i)} (1-p)^{S-n_0(i)}.$$

The right hand side (RHS) becomes

$$\begin{aligned}
RHS &= \textstyle\sum_{\{(j,U'):(j,i)\in V_R^2, U'=(U)_{ji}^+\}} \int_0^\infty C \left(\frac{\lambda p}{1-p}\right)^{m(j)} \frac{(1-p)^{S-n_0(j)}}{p^{1\{n_0(j)=0\}}} dF(u) \\
&= \textstyle\sum_{\{j:(j,i)\in V_R^2\}} C \left(\frac{\lambda p}{1-p}\right)^{m(j)} \frac{(1-p)^{S-n_0(j)}}{p^{1\{n_0(j)=0\}}} \\
&= \textstyle\sum_{\{j:(j,i)\in V_R^2\}} C \left(\frac{\lambda p}{1-p}\right)^{m(i)+1} \frac{(1-p)^{S-n_0(i)+n_{ij}}}{p^{1\{n_0(j)=0\}}} \\
&= \lambda C \left(\frac{\lambda p}{1-p}\right)^{m(i)} (1-p)^{S-n_0(i)} \left(\frac{p}{1-p}\right) \textstyle\sum_{\{j:(j,i)\in V_R^2\}} \frac{(1-p)^{n_{ij}}}{p^{1\{n_0(j)=0\}}} \\
&= \lambda C \left(\frac{\lambda p}{1-p}\right)^{m(i)} (1-p)^{S-n_0(i)} \left(\frac{p}{1-p}\right) \textstyle\sum_{n_{ij}=1}^{n_0(i)} \frac{(1-p)^{n_{ij}}}{p^{1\{n_0(j)=0\}}} \\
&= \lambda C \left(\frac{\lambda p}{1-p}\right)^{m(i)} (1-p)^{S-n_0(i)} \left(\frac{p}{1-p}\right) (\textstyle\sum_{k=1}^{n_0(i)-1} (1-p)^k + \frac{(1-p)^{n_0(i)}}{p}) \\
&= \lambda C \left(\frac{\lambda p}{1-p}\right)^{m(i)} (1-p)^{S-n_0(i)} \left(\frac{p}{1-p}\right) (\frac{1-p}{p}) \\
&= \lambda C \left(\frac{\lambda p}{1-p}\right)^{m(i)} (1-p)^{S-n_0(i)} \\
&= LHS.
\end{aligned}$$

Therefore, $\zeta_{(i,u)}$ is the stationary distribution of $Z(t)$.

∎

BIBLIOGRAPHY

[1] Baganha, M.P., Feeney and Sherbrooke Revisited, Draft, (1985).

[2] Feeney, G.J., Sherbrooke, C.C., The (S-1,S) Inventory Policy under Compound Poisson Demand, *Management Science,* Vol. 12, No. 5, Series A, Sciences., pp. 391-411 (Jan.,1966).

[3] Moinzadeh, K., Operating Characteristics of the (S-1,S) Inventory System with Paritial Backorders and Constant Resupply Times *Management Science*, Vol.35, No. 4., pp. 472-477 (Apr. 1989).

[4] Resnick, S.I., *Adventures in Stochastic Processes*, Birkhauser, Boston, (2005).

[5] Sevastyanov, B.A., An ergodic theory for Markov processes and its application to a telephone system with refusals. *Theory Prob. Appl.*, 2,104-112 (1957).

[6] Smeitink, E., A Note on "Operating Characteristics of the (S-1,S) Inventory System with Partial Backorders and Constant Resupply Times" *Management Science*, Vol. 36, No. 11, pp. 1413-1414 (Nov. 1990).

CHAPTER 3

AN EMERGENCY RESUPPLY NETWORK MODEL UNDER STUTTERING POISSON DEMAND

Jie Chen, Peter L. Jackson, John A. Muckstadt
School of Operations Research and Information Engineering,
Cornell University, Ithaca, NY, USA
jc562@cornell.edu, pj16@cornell.edu, jam61@cornell.edu

Abstract: We investigate the performance of employing an (S-1,S) inventory policy when demand follows a stuttering Poisson process and when demand is in excess of on-hand inventory, the excess demand is routed to an emergency order fulfillment system. This system contains a regional stocking location (RSL), which serves two types of facilities: a set of field service locations (FSL) and an emergency stocking location (ESL). The field service locations support technical service representatives who make visits to customer sites to repair equipment. We derive both exact and approximate expressions for the mean and variance of the number of units in emergency resupply. We also estimate the probability of zero units in emergency resupply. Simulation results reveal that the stationary distribution of the number of units in emergency resupply is well approximated by an atom at zero and a zero truncated negative binomial distribution. In particular, the approximation we develop is excellent in the upper tail which is the portion of the distribution used to determine the target inventory level for the emergency stocking location. Furthermore, the quality of the approximation appears to be insensitive to the actual form of the lead time distribution.

3.1 Introduction and Literature Review

3.1.1 Introduction

We investigate the effect of employing an (S-1,S) inventory policy when demand follows a stuttering Poisson process and when demand is in excess of on-hand inventory, the excess demand is routed to an emergency order fulfillment system. The system contains a regional stocking location (RSL), which serves two types of facilities: a set of field service locations (FSL) and an emergency stocking location (ESL). The field service locations support technical service representatives who make visits to customer sites to repair equipment. Figure 3.1 depicts one such system. The general problem addressed is setting stock levels to optimize performance. This paper is focused on estimating steady state performance measures of this system for use in optimization models.

We use the results from our companion paper, Chen *et al.*(2010) as the basis for our analysis. In that paper, we derive the exact stationary distribution for the number of units on order in the regular replenishment system for a single FSL when lead times are exponentially-distributed. In this paper, we develop both exact and approximate expressions for the mean and variance of the number of emergency orders outstanding in a system consisting of many field service locations and an emergency stocking location. We also estimate the probability of zero units in emergency resupply. Given these three statistics, we construct a distribution with an atom at zero combined with a zero-truncated negative binomial distribution to approximate the distribution of the number of units in emergency resupply. In a companion paper, we develop an optimization algorithm for setting stock levels in such a system.

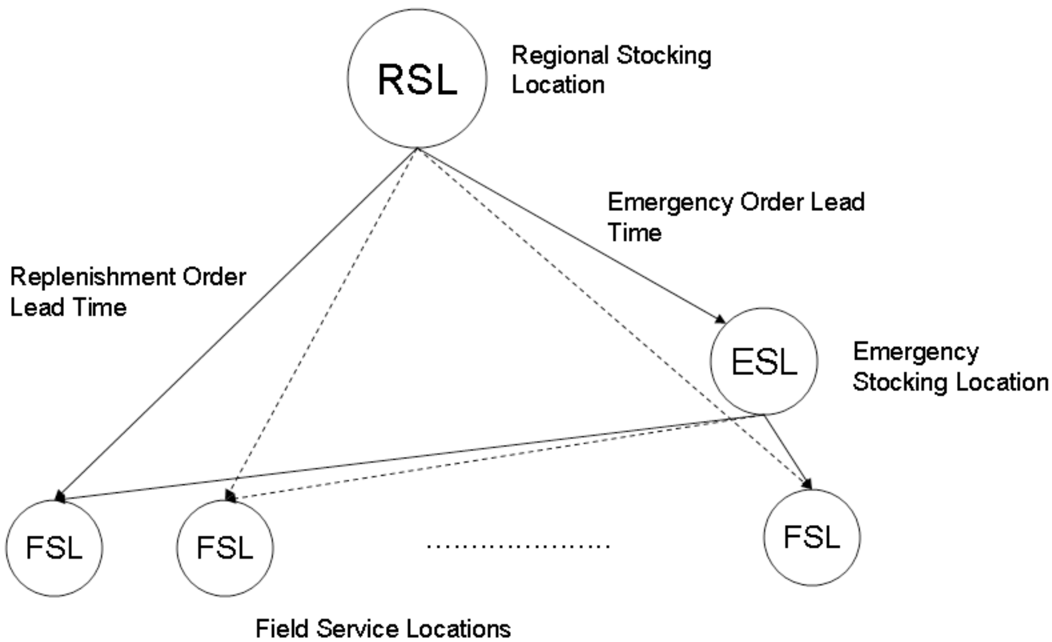

Figure 3.1: A System with Emergency Resupply

This paper is organized as follows. In the remainder of this section we review the literature. In Section 2 we develop the model for the complete fill case. By complete fill, we mean that a customer order is completely rejected (all units are lost) if there is insufficient stock on hand to fill it. The partial fill case (demand in excess of inventory on hand is lost), which is simpler to analyze, is covered in the appendix (3.F). Sections 3 and 4 provide exact and approximation methods, respectively for computing the first and second moments of the distribution of the number of units of different order sizes in resupply at the emergency stock location (ESL). In Section 5, we combine these results to focus on the total number of units in resupply at the ESL. We develop a method to approximate the distribution of the number of units in resupply at the ESL based on estimates of its mean and variance and the probability of zero units in emergency resupply. We also investigate the quality of the approximation using simulation. The quality of the approximation when used for non-exponential lead time distributions is also considered. Concluding comments are found in Section 6. An experimental results and proofs of all theorems are found in the appendix.

3.1.2 Review of Literature

For the single-location inventory system with multiple shipment modes, early papers by Barankin (1961), Daniel (1962), Neuts (1964), Bulinskaya (1964) and Veinott (1966) study optimal ordering policies for periodic review inventory systems with a nominal lead time of one period and an emergency lead time of zero periods. Fukuda (1964) and Wright (1969) allow for multi-period nominal lead times; but, the lead time of the emergency order is always one period shorter than that of nominal order lead time. Rosenshine and Obee (1976) examine the effectiveness of a standing order inventory system (fixed-size order arrives at the beginning of each period), allowing a zero lead time emergency order if large shortage occurs and sell-offs if inventory exceeds storage capacity. Whittemore and Saunders (1977) derive the optimal policy for an infinite horizon inventory system allowing two types of orders with multi-period nominal lead times. Blumenfeld *et al.* (1985) analyze the trade-off between safety stock (inventory cost) and emergency order penalties assuming that the emergency order is placed only once in a review period, when a shortage occurs, and arrives immediately with a sufficient amount to meet any demands before the next regular shipment.

Instead of placing both regular and emergency orders in a review period, the papers of Gross and Soriano (1972) and Chiang and Gutierrez (1996) analyze the decision rule for choosing the shipment mode for the whole resupply order which needs to be placed. Chiang and Gutierrez (1996) consider emergency resupply lead times which are less than one review period in length. In a sequel, Chiang and Gutierrez (1998) study a different inventory system with emergency orders placed on a continuous basis while regular orders are placed periodically. Tagaras and Vlachos (2001) propose a simple approximate model for a similar type of inventory system assuming that during each cycle period, the emergency orders may be issued only once at specific review epochs. Teunter and Vlachos (2001) allow for emergency orders to arrive some fixed number of time units before a regular resupply order is due. This could be considered as a generalization of the emergency order conditions of the former two papers. Zhang (1996), Feng *et al.*(2005, 2006) discuss the optimality of the base-stock policy with multiple delivery modes in different situations.

For periodic review multi-stage supply chain systems with emergency orders, the paper of Lawson and Porteus (2000) studies a serial supply chain system allowing expedited shipment with zero lead time between any adjacent echelons, where the regular lead time is always one period. Muharremoğlu and Tsitsiklis (2003) extend the model of Lawson and Porteus (2000) by allowing complete lead time flexibility with a more general cost structure for the expedited shipping units from any stage in the system to any downstream stage. Huggins and Olsen (2003, 2005) examine optimal policies for two-stage supply chain system under both centralized and decentralized control where expediting can be used to satisfy unmet demand in each period.

For continuous review inventory models with emergency orders, Allen and

D'Esopo (1968) propose three operational parameters: reorder point, order quantity, and expediting level. They propose that whenever the inventory level drops to the expediting level, an outstanding order will be expedited and delivered after a constant period which is shorter than the constant regular lead time. Muckstadt and Thomas (1980) analyze multi-item multi-echelon inventory systems with $(S-1, S)$ inventory policy and Poisson demand considering emergency orders as lost sales. Moinzadeh and Nahmias (1988) propose an extension of the (Q, R) policy with different reorder points and reorder sizes for regular and emergency orders. In deriving cost expressions, they assume that there is never more than one outstanding order of each type. Assuming Poisson demand, Moinzadeh and Schmidt (1991) consider a single location (S-1,S) inventory system with two options (regular and emergency) resupply. Emergency orders arrive in a shorter time but at a higher cost. Assuming that lead times are known and constant, they propose an order policy incorporating the age of the outstanding orders. Moinzadeh and Aggarwal (1997) adopt the same class of policies but extend the results to a multi-echelon inventory system. Considering the standard (Q, R) policy for regular replenishment orders with constant lead time, Johansen and Thorstenson (1998) let the emergency orders (with short, constant lead times) depend also on the remaining delivery time for a regular order. Chiang (2002) proposes two single-location, single-item policies when expediting is allowed. One policy modifies the Allen and D'Esopo (1968) policy by adding a threshold time point, which is the last point when expediting is allowed, while the other policy allows an expediting decision to be made at some point during the lead time. The problem of using multiple suppliers efficiently is reviewed by Minner (2003). Axsäter (2005) considers a single-echelon continuous review inventory system facing compound Poisson demand. The system gets normal replenishments according to a standard (Q, R) policy and emergency replenishments in critical situations. Axsäter suggests a heuristic decision rule for triggering emergency orders, which minimizes the expected costs under the assumption that there is only at most one emergency replenishment outstanding at any time.

An extensive literature is also available studying *lateral transshipments* among locations. Many papers examine the effect of employing decision rules for making lateral transshipments. Examples of this type are Lee (1987), Axsäter (1990), Dada (1992), Alfredsson and Verrijdt (1999), Grahovac and Chakravarty (2001), Kukreja *et al.* (2001), Sherbrooke (1992), Muckstadt (2005), Vidgren (2005), Axsäter (2006) and Vliegen (2009). Other papers present methods for optimizing the decisions concerning lateral transshipments. Examples are Das (1975), Robinson (1990), Tagaras and Cohen (1992), Archibald *et al.* (1997), Rudi *et al.* (2001), Minner *et al.* (2003), Wong *et al.* (2006), Olsson (2009), Kranenburg and van Houtum (2009), Wijk *et al.* (2009) and Reijnen *et al.*(2009). Paterson *et al.* (2009) provide an up-to-date review of the inventory models with lateral transshipment.

The system considered in this paper differs from others in the literature in that it explicitly considers a stocking location that is dedicated to satisfying emergency orders, and attempts to estimate the distribution of outstanding orders for this type of location. For the purposes of analysis, we assume continuous review $(S-1, S)$ stocking policy Stuttering Poisson demand and exponentially distributed lead times. We also investigate constant lead times in an empirical study.

3.2 Emergency Orders with Compound Poisson Demand and Exponential Lead Times

In this paper, we extend the lost sales model in Chen *et al.*(2010) to an emergency resupply network system in which demand processes at field stock locations are stuttering Poisson processes. We assume that once a customer demand exceeds the inventory on hand, an emergency order occurs equal to the order size of the customer. If we treat the emergency order as a 'lost sale', the replenishment process to the FSLs from the RSL behaves as a lost sales model with complete fill. We could also analyze the model with partial fill. Since both cases are addressed using the same approach, we present the analysis only for the complete fill case. The partial fill case is treated in the appendix.

In the emergency order fullfillment systems that we have observed in practice, the demand processes at the FSL's exhibit a high variance-to-mean ratio. It is not accurate to model such processes simply as Poisson processes. Rather, we model demand as a stuttering Poisson process and propose statistically fitting the two parameters of this process using historical data.

We assume the ESL is geographically close to all of the FSLs assigned to it (there are many ESL-FSL sets served by the RSL). The ESL manages inventory by following an $(S-1, S)$ policy as well, so every emergency order is immediately passed back to the RSL as a replenishment order for the ESL. The replenishment order lead times for the FSL's are roughly the same as the replenishment order lead time for the ESL. These times are typically measured in days. The time to order from the ESL and to deliver an emergency order to the FSL is short, typically measured in hours.

For the purpose of analysis we make a number of simplifying assumptions. First, we assume that the RSL has infinite stock on hand and the demand processes at the different FSL's are mutually independent stuttering Poisson processes. The parameters of these processes may differ by location. We also assume that emergency orders that cannot be satisfied from stock at the ESL are backlogged at the ESL. Furthermore, we assume that the lead time from the ESL to the FSL is short enough to make emergency resupply from the ESL desirable. Since we track (and penalize) backorders only at the ESL, the lead time between the ESL and the FSL is not considered in our analysis. In a companion paper we construct cost and customer service measures that depend on the steady state distributions of the number of units in regular replenishment from the RSL to the FSL (i.e. in regular resupply) and the number of units in regular replenishment from the RSL to the ESL (i.e. in emergency resupply).

As mentioned, the RSL-FSL system can be modeled as a lost sales system. Chen *et al.*(2009) show how to compute the steady state distribution of the number of units in regular resupply at the FSLs. In this paper, we focus on estimating the steady state distribution of the number of units in emergency resupply. We focus first on a single FSL and examine the number of units from emergency orders that will be in resupply (RSL to ESL) corresponding to that one FSL. We derive the mean and variance of the number of units in emergency resupply corresponding to demand occurring at a single FSL. We also estimate the probability of zero units in emergency resupply. Subsequently, we aggregate these statistics over multiple independent FSL emergency order streams.

The replenishment order lead times from the RSL to the ESL are assumed to be independent and exponentially distributed with mean: $\tau_E = 1/\mu$. This assumption is relaxed in the simulation experiments of Section 5.2.3.

Let $Y_k(t)$ denote the number of replenishment orders of size k outstanding at the ESL at time t from a single FSL, for $k = 1, 2, \ldots$. Let Y_k denote the random variable with the limiting distribution of $Y_k(t)$ as $t \to \infty$. Let $Y(t)$ denote the total number of replenishment orders in emergency resupply. That is, $Y(t) = \sum_{k=1}^{\infty} Y_k(t)$. Let Y denote the random variable with the limiting distribution of $Y(t)$ as $t \to \infty$.

Let Λ_h^k denote the hth moment of the replenishment orders of size k outstanding at the ESL, that is,

$$\Lambda_h^k \equiv E[Y_k^h],\ h = 1, 2, 3.$$

Let $Z(t)$ denote the total number of units in resupply for the ESL at time t. That is, $Z(t) = \sum_{k=1}^{\infty} kY_k(t)$. Let Z denote the random variable corresponding to the limiting distribution of $Z(t)$ as $t \to \infty$. Under the assumptions of an model these limiting distributions are well-defined in the following sense.

Proposition 6 *For any fixed target stock level S, $\Lambda_h^k = E(Y_k^h) < \infty$ for $h = 1, 2, 3$. The stationary distribution Z of the number of the emergency ordered units has*

$$\sum_{k=1}^{K} kY_k \xrightarrow{L^2} Z, \text{as} K \to \infty.$$

Hence $E(Z) = \sum_{k=1}^{\infty} kE(Y_k)$ and

$$Var(Z) = \sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} k_1 k_2 Cov(Y_{k_1}, Y_{k_2}).$$

Similarly, the stationary distribution Y of the number of the emergency orders has $\sum_{k=1}^{K} Y_k \xrightarrow{L^2} Y,$ as $K \to \infty$. Thus $E(Y) = \sum_{k=1}^{\infty} E(Y_k)$ and

$$Var(Y) = \sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} Cov(Y_{k_1}, Y_{k_2}).$$

The goal of this paper is to develop an approximation for the stationary distribution of Z, the number of units in emergency resupply. By Proposition 6,

$$Var(Y) = \sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} Cov(Y_{k_1}, Y_{k_2}) \approx \sum_{k_1=1}^{K}\sum_{k_2=1}^{K} Cov(Y_{k_1}, Y_{k_2}),$$

and

$$Var(Z) = \sum_{k_1=1}^{\infty}\sum_{k_2=1}^{\infty} k_1 k_2 Cov(Y_{k_1}, Y_{k_2}) \approx \sum_{k_1=1}^{K}\sum_{k_2=1}^{K} k_1 k_2 Cov(Y_{k_1}, Y_{k_2}),$$

for some large K. Hence, we estimate the mean and variance of Y and Z by estimating the mean, variance, and mutual covariances of a finite number of the Y_k's. Furthermore, we use the mean and variance of Y to estimate $P(Y = 0)$ which is also the probability of no units in emergency resupply, that is, $P(Z = 0)$. From these statistics we construct an approximate distribution for Z.

3.3 The First and Second Moments of the Number of Emergency Orders

In this section, we develop an exact method for determining the first two moments of Y, the number of emergency orders, of size k. First, let us review the continuous time model developed in Chen., *et al.*(2009) in which demand arrives according to a stationary compound Poisson process. Let λ denote the rate of arrivals of customers and let X denote the order size random variable which is positive and integer-valued. Let $p_k \equiv P\{X = k\}$ and let $\overline{P}_k \equiv P\{X > k\}$ for all $k = 0, 1, 2, ...$ We assume at least one unit is ordered for each customer arrival: $p_0 = 0$ and $\overline{P}_0 = 1$, although the results are easily generalized to allow for zero-sized orders. For the special case of the so-called stuttering Poisson process, the order size distribution is geometric. Let p denote the probability of a unit-sized order under the geometric distribution: $p_1 = p$. In this case, for all $k = 1, 2, ...,$ $p_k = p(1-p)^{k-1}, \overline{P}_k = (1-p)^k$.

Let I_t denote the inventory on hand at time t, $t \geq 0$, a non-negative, integer-valued random variable. In the complete fill case, if the customer order size X_t is larger than I_t, the order is rejected. Since the system is managed according to an $(S-1, S)$ policy, when a customer order with size k is accepted, we place a replenishment order with the same size k.

Finally, we assume that lead times for regular replenishment orders (the RSL to the FSL) are independent, exponentially-distributed random variables with rate μ_F with mean $\tau_F = 1/\mu_F$.

Let N_{kt} denote the number of regular replenishment orders of size k outstanding at time t at an FSL, for $k = 1, 2, ..., S$, and let $N_t = (N_{kt})_{k=1}^{S}$ denote the vector of outstanding regular replenishment orders. Given our assumptions of lost sales, complete fills, and an $(S-1, S)$ policy, it follows that

$$I_t + \sum_{k=1}^{S} k N_{kt} = S.$$

The stochastic process $N = \{N_t, t \geq 0\}$ is a finite-state, time-homogeneous Markov process. Let V index the state space of the underlying Markov chain. That is, we assume the existence of a one-to-one mapping from V to the set of all possible vectors of outstanding replenishment orders. For each $i \in V$, we denote the mapping by $n(i) = (n_1(i), n_2(i), ..., n_S(i))$ where $n_k(i) \in \{0, 1, ..., \lfloor S/k \rfloor\}$ forall$k = 1, ..., S$, and $\sum_{k=1}^{S} k n_k(i) \leq S$. Furthermore, the implied number of units on hand is

$$n_0(i) \equiv S - \sum_{k=1}^{S} k n_k(i).$$

Let $m(i) \equiv \sum_{k=1}^{S} n_k(i)$, be the total number of outstanding orders in state i. The infinitesimal generator for the Markov process N at a FSL is given by

$$A_{ij} \equiv \begin{cases} n_{k_{ij}}(i)\mu_F & \text{if}(i,j) \in V_R^2, \\ \lambda p_{k_{ij}} & \text{if}(i,j) \in V_C^2, \\ -(m(i)\mu_F + \lambda(1-\overline{P}_{n_0(i)}))1_{\{n_0(i)\neq 0\}} & \text{if}\, j = i, \\ 0 & \text{otherwise,} \end{cases} \tag{3.1}$$

where $1_{\{E\}}$ is the indicator function of condition E ($1_{\{E\}} = 1$ if E happens and $= 0$ otherwise)

Recall that the transition $(i, j) \in V_R^2$ means a replenishment of size k_{ij} arrived at this FSL. In this case, the transition rate is $n_{k_{ij}}(i)\mu_F$. The condition $(i, j) \in V_C^2$ means an order of size k_{ij} arrives and could be satisfied completely. So the transition rate is $\lambda p_{k_{ij}}$. For any other (i, j), $j \neq i$, there is no single step transition between them, so the transition rate is zero. Finally, for $j = i$:

$$A_{ii} = -\sum_{j\in V, j\neq i} A_{ij} = -(m(i)\mu_F + \lambda(1-\overline{P}_{n_0(i)})1_{\{n_0(i)\neq 0\}}).$$

As shown in Chen *et al.*(2009), the stationary distribution of this process, N, is given by ξ_i for any state $i \in V$, where

$$\xi_i \equiv \frac{\left(\frac{\lambda p}{\mu_F(1-p)}\right)^{m(i)}}{\prod_{k=1}^{S}(n_k(i)!)\, G(S)} (1-p)^{S-n_0(i)}, \tag{3.2}$$

and $G(S)$ is the normalizing constant. Let $f_{NB}(x; m, p)$ denote the negative binomial probability of x failures before the m^{th} success where the probability of success is p. For the special case of $m = 0$, we use $f_{NB}(x; 0, p) = 1_{\{x=0\}}$. Then,

$$G(S) = \sum_{s=0}^{S}\sum_{m=0}^{s} \frac{(\frac{\lambda}{\mu_F})^m}{m!} f_{\text{NB}}(s-m; m, p). \tag{3.3}$$

Furthermore, the stationary distribution of the number of units on order, π_s ($s = 0, 1, \ldots, S$), is given by:

$$\pi_s = \frac{\sum_{m=0}^{s}(\frac{\lambda}{\mu_F})^m \frac{f_{NB}(s-m;m,p)}{m!}}{G(S)}. \tag{3.4}$$

Equation (3.4) is equivalent to the classical lost sales formula in Feeney and Sherbrooke (1966); but the derivation in Chen *et al.*(2009) is new.

We next create an enlarged state space that will provide us with the capability to determine the desired statistics for the emergency network system. Define the k^{th} system to consist of all orders, by order size, outstanding at the FSL together with $Y_k(t)$, the number of orders of exactly size k outstanding at the ESL from this FSL at time t. This can be modeled as a continuous time Markov Chain. For each $k = 1, 2, \ldots$, we extend the state space V of the FSL system into $V \times N^+$, which is the state space of the k^{th} system. Recall that $n(i) = (n_1(i), n_2(i), \ldots, n_S(i))$ for $i \in V$. Let $N_t^k \equiv (N_{1t}, N_{2t}, \ldots, N_{St}, Y_k(t)) = (N_t, Y_k(t))$ denote the vector of outstanding replenishment and emergency orders of size k at time t. By extending the notation introduced in Chen *et al.* (2010), for any state $(i, y) \in V \times N^+$, we have $N_t^k(i, y) = (N_t = n(i), Y_k(t) = y) = (n_1(i), n_2(i), \ldots, n_S(i), y)$.

Let $(\psi_{i,y}^k)$ denote the steady state distribution for the k^{th} system. Observe that $\sum_{y=0}^{\infty} \psi_{i,y}^k$ is the probability that the state of the FSL replenishment orders is i, that is,

$$\xi_i = \sum_{y=0}^{\infty} \psi_{i,y}^k.$$

In addition, let $\psi_{y|i}^k$ denote the steady state probability that y orders of size k are in resupply at the ESL conditioned on state i at the FSL. Then

$$\psi_{y|i}^k \xi_i = \psi_{i,y}^k.$$

Let $\Lambda_{h,i}^k$ denote the hth moment of the outstanding replenishment orders of size k at the ESL when the replenishment order state is $i \in V$ at the FSL. We require three moments:

$$\Lambda_{h,i}^k \equiv E[Y_k^h 1_{\{N=n(i)\}}] = \sum_{y=0}^{\infty} y^h \psi_{i,y},\ h = 1, 2, 3.$$

Therefore

$$\Lambda_h^k = \sum_{i \in V} \Lambda_{h,i}^k,\ h = 1, 2, 3.$$

.

Define $\rho_k(i) \equiv 1_{\{k>n_0(i)\}} \lambda p(1-p)^{k-1}$ as the arrival rate for $Y_k(t)$ conditioned on the state of the FSL replenishment process.

Proposition 7 *For any fixed k, the moments $\Lambda_{1,i}^k$ satisfy the following equations:*

$$0 = \sum_{j \in V} A_{j,i} \Lambda_{1,j}^k - \mu \Lambda_{1,i}^k + \xi_i \rho_k(i), \text{for} i \in V. \tag{3.5}$$

That is, the row vector $\overrightarrow{\Lambda_{1\cdot}^k} \equiv (\Lambda_{1,i}^k)_{1\times|V|}$ has elements given by

$$\overrightarrow{\Lambda_{1\cdot}^k} = -(\xi_i \rho_k(i))_{1\times|V|} \times (A - \mu I_{|V|\times|V|})^{-1}, \tag{3.6}$$

where $A = (A_{ij})$, $|V|$ is the number of states of V and $I_{|V|\times|V|}$ is the $|V|$-dimensional identity matrix.

In addition,

$$\Lambda_1^k = \frac{\sum_{i \in V} \xi_i \rho_k(i)}{\mu}. \tag{3.7}$$

and

$$\Lambda_2^k = \sum_{i \in V} [\frac{\Lambda_{1,i}^k(\mu + 2\rho_k(i)) + \xi_i \rho_k(i)}{2\mu}]. \tag{3.8}$$

When S is small enough, say 15 or less, then $(A - \mu I)^{-1}$ is readily computable and we can solve equations (3.5) for $\Lambda^k_{1,j}$ and then solve (3.7) for Λ^k_1. Further we get Λ^k_2 by (3.8) for any size k. However, $|V|$ grows rapidly as S increases so this exact approach is computationally intractable when the demand over a lead time is substantial and shortages are to be avoided. For a more general approach, we use an approximation to decrease the dimension of the state space from $|V|$ to $S + 1$.

3.4 The Mean and Second Moment Approximation of the Number of the Emergency Orders

When the approach described in the previous section is computationally intractable, we need an alternative method to estimate the moments. To do this, we first construct a Markov Chain to approximate the steady state distribution of the number of units on order at the FSL. We then show how to simplify the formulas in Proposition 7 using this approximating Markov chain. Finally, we define a similar method for estimating moments of a system with pairs of order sizes outstanding.

3.4.1 Markov Chain Approximation for the Number of Units on Order at the FSL

We define a new continuous time Markov chain on the number of units on order at the FSL, where has state space $\mathcal{S} \equiv \{0, 1, ..., S\}$. We seek to define the infinitesimal generator, $Q_{(S+1)\times(S+1)}$, for this Markov chain so that it has the same stationary distribution, denoted $(\tilde{\pi}_s)$, on this reduced state space, as given by the stationary distribution (π_s) for the complete fill case as given in equation (3.4). A similar approach is possible for the partial fill case.

Consider two states s and s' in $\mathcal{S}$ such that $s \neq s'$, and consider a single step transition from one state, s, to the other, s'. If $s < s'$, then the transition corresponds to a customer arrival with order size exactly equal to $s' - s$. The rate of such transitions is $Q_{s,s'} = \lambda p (1-p)^{s'-s-1}$.

On the other hand, if $s > s'$, then the transition corresponds to an order delivery. The rate of such transitions cannot be determined without knowing the number of orders outstanding. We condition, therefore, on m, the number of orders outstanding. In the true lost sales Markov chain, the conditional probability of m orders outstanding given s units on order is $\xi_{m,s}/\pi_s$, where π_s is given by (3.4) and, in the complete fill case, $\xi_{m,s}$ is given by

$$\xi_{m,s} = \sum_{\substack{i \in V \\ S-n_0(i)s \\ m(i)=m}} \xi_i = \left(\frac{\lambda}{\mu_F}\right)^m \frac{f_{NB}(s-m; m, p)}{m! G(S)}, \tag{3.9}$$

where $G(S)$ is the same normalizer as stated in (3.3).

Given m, s, and s', when $s > s'$, the transition rate is still indeterminate because it depends on the relative likelihood of an order of size $s - s'$ being delivered. Let $g(m, s, s')$ denote the following approximation to this likelihood:

$$g(m,s,s') \equiv \begin{cases} \frac{\binom{s'-1}{m-2}}{\binom{s-1}{m-1}} & \text{if} m > 1, s > s' > 0, \\ 1 & \text{if} m = 1, s > s' = 0, \\ 0 & \text{otherwise.} \end{cases}$$

A combinatorial interpretation of this approximation is as follows. Suppose we know only that out of m orders outstanding, with s units on order, a single delivery is made of size $s-s'$. If $m=1$, then it must be the case that $s'=0$ and there is only one way in which this can happen (i.e. $g(1,s,0)=1$). On the other hand, if $m>1$, then there are $\binom{s-1}{m-1}$ ways in which there can be m remaining orders with a total of s units as shown in Chen,*et al.* (2010). Of all these ways, we are interested only in those ways that could lead to a single delivery of size $s-s'$. There are $\binom{s'-1}{m-2}$ ways in which there can be $m-1$ remaining orders with a total of s' units. Each one of these ways matches exactly one of the ways of grouping s units into m orders, i.e. by adding a single order of size $s-s'$. Assuming that each of these ways is equally likely, the ratio $\frac{\binom{s'-1}{m-2}}{\binom{s-1}{m-1}}$ expresses the likelihood that we will see a delivery of size $s-s'$ given that we see a delivery when m orders are outstanding and s units are on order.

Assembling these components for the case $s>s'$, let $\gamma(s,s')$ denote the approximate rate of transitions from state s to state s':

$$\gamma(s,s') \equiv \sum_{m=1}^{s'+1} m\mu_F g(m,s,s')\frac{\xi_{m,s}}{\pi_s}.$$

Lemma 4 *For any positive integers s,m and $s>m$,*

$$\sum_{l=1}^{s-m+1} g(m,s,s-l) = 1.$$

We are led to defining Q, the infinitesimal generator of the approximating Markov chain, as

$$Q(s,s') \equiv \begin{cases} \lambda p\,(1-p)^{s'-s-1} & 0 \le s < s' \le S, \\ \gamma(s,s') & 0 \le s' < s \le S, \\ -\left(\Gamma(s) + \lambda \sum_{l=1}^{S-s} p(1-p)^{l-1} 1_{\{s \neq S\}}\right) & s = s' \in \{0,1,2,...,S\}, \\ 0 & \text{otherwise}, \end{cases}$$

where $\Gamma(s) \equiv \sum_{l=1}^{s} \gamma(s, s-l) 1_{\{s \neq 0\}}$.

By construction, this Markov chain's stationary distribution, denoted $\{\tilde{\pi}_s, s = 0, 1, 2, \ldots, S\}$, satisfies the following balance equations:

$$[\lambda \sum_{l=1}^{S-s} p(1-p)^{l-1} + \Gamma(s)]\tilde{\pi}_s = \lambda \sum_{l=1}^{s} p(1-p)^{l-1} \tilde{\pi}_{s-l} + \sum_{l=1}^{S-s} \gamma(s+l, s)\tilde{\pi}_{s+l}, \text{for} s = 1, 2, \ldots, \tag{3.10}$$

and $\sum_{s=0}^{S} \tilde{\pi}_s = 1$.

Proposition 8 *In the complete fill case, the stationary distribution π_s, given by (3.4), also satisfies balance equation (3.10). That is, $\tilde{\pi}_s = \pi_s$ for $s = 0, 1, 2, \ldots, S$.*

The generators A and Q are defined on different state spaces. We relate these two generators in the following lemma.

Lemma 5 *For each $i \in V$ and $n(i) = (n_1(i), n_2(i), \ldots, n_s(i))$ at the FSL with a stuttering Poisson demand process (complete fill case), for fixed $s = 0, 1, \ldots, S$,*

$$\sum_{i:n_0(i)=S-s} \xi_i A_{i,i} = \pi_s Q(s, s), \tag{3.11}$$

and, for fixed $d \in \{0, 1, \ldots, S\}$ and $d \neq s$,

$$\sum_{i:n_0(i)=S-d} [\sum_{j:n_0(j)=S-s} \xi_i A_{i,j}] = \pi_d Q(d, s). \tag{3.12}$$

In summary, we have defined the generator Q of a Markov chain on the number of units on order at the FSL such that it yields the same steady state distribution as the true lost sales model for the complete fill case. We also derived an equivalence (Lemma 5) that is used in the next subsection to collapse the state space.

3.4.2 The Mean and Second Moment Approximation

By collapsing the state space, we can use Q, the generator of the approximating Markov chain, to simplify the formula for the second moment of the replenishment orders of size k (3.8).

Recall that $\psi^k_{i,y}$ is the steady state distribution of the k^{th} system and $\psi^k_{y|i}$ is the steady state distribution of y orders of size k at the ESL conditioned on state i at the FSL. Recall also that $\rho_k(i)$ is the rate at which emergency orders of size k are generated, conditioned on the state of the FSL replenishment process.

We are also interested in moments of emergency orders joint with the number of units on order. Let $\tilde{\Lambda}^k_{1,s}$ denote the expected number of the replenishment orders of size k at the ESL when there are s units on-order at the FSL. That is,

$$\tilde{\Lambda}^k_{1,s} = \sum_{i:n_0(i)=S-s} \Lambda^k_{1,i} = E[\sum_{i:n_0(i)=S-s} Y_k 1_{\{N=n(i)\}}],\ s \in \{0, 1, \ldots, S\}.$$

Observe that $\Lambda^k_{1,i}$ is defined on the state space V, and $\tilde{\Lambda}^k_{1,s}$ is defined on the state space $\{0, 1, \ldots, S\}$. We have the following relationship:

$$\Lambda^k_1 \equiv \sum_{s=0}^{S} \tilde{\Lambda}^k_{1,s} = \sum_{i \in V} \Lambda^k_{1,i}.$$

By definition, $\rho_k(i)$ depends only on the number of units on hand, $n_0(i)$, so when collapsing the state space we will use $\tilde{\rho}_k(s)$ to mean $\rho_k(i)$ for any state i with $S - n_0(i) = s$. That is,

$$\tilde{\rho}_k(s) = 1_{\{k>S-s\}} \lambda p(1-p)^{k-1}.$$

Rewriting (3.7) in terms of the reduced state space, we have:

$$\Lambda^l_1 = \frac{\sum_{s=1}^{S} \pi_s \tilde{\rho}_k(s)}{\mu}, \tag{3.13}$$

which follows easily from Little's Law.

Suppose that the conditional stationary distribution of the number of emergency orders of size k given the state of the FSL $\psi^k_{y|i}$ depends only on the quantity of on hand inventory at the FSL. Although this is not true in general, it is a useful approximation.

Proposition 9 *Suppose $\psi^k_{y|i} = \psi^k_{y|i'}$, $\forall i, i' \in \{j \in V : n_0(j) \equiv n_0(j')\}$. Then for $d = 0, 1, \ldots, S$, $\tilde{\Lambda}^k_{1,d}$ will satisfy*

$$0 \equiv \sum_{d=0}^{S} \tilde{\Lambda}^k_{1,d} Q(d,s) - \mu \tilde{\Lambda}^k_{1,s} + \pi_s \tilde{\rho}_k(s),$$

for each $s = 0, 1, \ldots, S$. Consequently their values can be determined by solving

$$(\tilde{\Lambda}^k_{1,s})_{1 \times S+1} = -(\pi_s \tilde{\rho}_k(s))_{1 \times S+1} (Q - \mu I_{S+1 \times S+1})^{-1},$$

where $I_{S+1 \times S+1}$ is the $S+1$-dimensional identity matrix. Furthermore,

$$\Lambda^k_2 = \sum_{d=0}^{S} [\frac{\tilde{\Lambda}^k_{1,d}(\mu + 2\tilde{\rho}_k(d)) + \pi_d \tilde{\rho}_k(d)}{2\mu}]. \tag{3.14}$$

Through collapsing the state space, we have an exact and computable expression for Λ^k_1, namely (3.13), and an approximate but computable expression for Λ^k_2, namely (3.14). For an example with $k = 1$ and $S = 2$, we compute both the exact (3.8) and approximate (3.14) values for Λ^k_2. We find that the approximation understates the true value by a small amount ($\sim 1.5\%$). Apparently, the assumption in Proposition 9 concerning the conditional stationary distribution of the number of emergency orders does not hold in general. However, in the larger context of approximating the moments of the total number of units in emergency resupply, Section 3.B in the appendix shows that the approximation is quite good.

3.4.3 Extension to a System Considering Two Different Order Sizes

Let the (k_1, k_2)-system consist of the orders and order sizes outstanding at the FSL and the number of orders of size k_1 and k_2 outstanding at the ESL. This also can be modeled as a continuous time Markov Chain.

The (k_1, k_2)-system has the following relationship with the $k_{1^{st}}$ system and the $k_{2^{nd}}$ system. For k_1, k_2 fixed and $k_1 \neq k_2$, let $Y_{k_1k_2,t}$ denote the number of the emergency orders of size k_1 or k_2 at time t. Thus $Y_{k_1k_2,t} = Y_{k_1t} + Y_{k_2t}$. Let $Y_{k_1k_2}$ denote the stationary distribution of $Y_{k_1k_2,t}$; it follows that $Y_{k_1k_2} = Y_{k_1} + Y_{k_2}$.

An exact analysis for deriving the moments of Y_{k_1,k_2} is possible using the approach of Section 3. In the interest of space, we present only the approximation approach. We focus on the collapsed state conditioned on s, the number of units on order at the FSL, $s \in \{0, 1, \ldots, S\}$. Let $\tilde{\rho}_{k_1k_2}(s)$ be the arrival rate for the emergency orders in this system space. Accordingly,

$$\tilde{\rho}_{k_1k_2}(s) = \lambda p[1_{\{k_1 > S-s\}}(1-p)^{k_1-1} + 1_{\{k_2 > S-s\}}(1-p)^{k_2-1}] = \tilde{\rho}_{k_1}(s) + \tilde{\rho}_{k_2}(s).$$

Define the expected number of the replenishment orders of the (k_1, k_2)-system at the ESL given uncollapsed state $i \in V$ at the FSL as

$$\Lambda_{1,i}^{k_1k_2} = E(Y_{k_1k_2} 1_{\{N=n(i)\}}),$$

and define the moments of Y_{k_1,k_2} as

$$\Lambda_h^{k_1k_2} \equiv E[Y_{k_1k_2}^h], \text{for} h = 1, 2.$$

Now define the expected number of the emergency replenishment orders of size k_1 and k_2 at the ESL given the collapsed state of s units on-order at the FSL as

$$\tilde{\Lambda}_{1,s}^{k_1 k_2} = \sum_{i:n_0(i)=S-s} E(Y_{k_1 k_2} 1_{\{N=n(i)\}}).$$

Consequently, the mean of the total number of replenishment orders given k_1 and k_2 at the ESL is given by

$$\Lambda_1^{k_1 k_2} = \sum_{s=0}^{S} \tilde{\Lambda}_{1,s}^{k_1 k_2}.$$

Corollary 3 *Given $\tilde{\Lambda}_{1,d}^{k_1}$ and $\tilde{\Lambda}_{1,d}^{k_2}$, for each $d = 0, 1, \ldots, S$, the first moment of Y_{k_1,k_2} is given by*

$$\Lambda_1^{k_1 k_2} = \sum_{d=0}^{S} (\tilde{\Lambda}_{1,d}^{k_1} + \tilde{\Lambda}_{1,d}^{k_2}). \tag{3.15}$$

The second moment is given by

$$\Lambda_2^{k_1 k_2} = \frac{\sum_{d=0}^{S} (\tilde{\Lambda}_{1,d}^{k_1} + \tilde{\Lambda}_{1,d}^{k_2})[\mu + 2(\rho_{k_1}(d) + \rho_{k_2}(d))] + \pi_d(\rho_{k_1}(d) + \rho_{k_2}(d))}{2\mu}. \tag{3.16}$$

In summary, we have developed a technique to approximate the first two moments of the number of emergency orders for each order size individually, (3.13) and (3.14), respectively, and for pairs of different order sizes, (3.15) and (3.16).

3.4.4 Statistics for the Number of Units in Emergency Resupply

We now turn to approximating the moments of Z and the probability of zero units in the emergency resupply, that is $P(Z = 0)$.

From Corollary 3, for $\forall k_1, k_2$ and $k_1 \neq k_2$:

$$\begin{aligned}
2Cov(Y_{k_1}, Y_{k_2}) &= Var(Y_{k_1k_2}) - Var(Y_{k_1}) - Var(Y_{k_2}) \\
&= (\Lambda_2^{k_1k_2} - (\Lambda_1^{k_1} + \Lambda_1^{k_2})^2) - (\Lambda_2^{k_1} - (\Lambda_1^{k_1})^2) - (\Lambda_2^{k_2} - (\Lambda_1^{k_2})^2) \\
&= \Lambda_2^{k_1k_2} - \Lambda_2^{k_1} - \Lambda_2^{k_2} - 2\Lambda_1^{k_1}\Lambda_1^{k_2}.
\end{aligned}$$

Therefore, $Var(Z)$ can be approximated by

$$Var(Z) \approx \sum_{k_1=1}^{K}\sum_{k_2=1}^{K} k_1k_2Cov(Y_{k_1}, Y_{k_2}) = \sum_{k=1}^{K} k^2(\Lambda_2^k - (\Lambda_1^k)^2) + \sum_{k_1<k_2\leq K} k_1k_2(\Lambda_2^{k_1k_2} - \Lambda_2^{k_1} - \Lambda_2^{k_2} - 2\Lambda_1^{k_1}\Lambda_1^{k_2}), \tag{3.17}$$

for some K large enough. Similarly,

$$E(Z) \approx \sum_{k=1}^{K} k\Lambda_1^k. \tag{3.18}$$

The probability of no units in emergency resupply, $P(Z = 0)$, is the same as the probability of no orders in emergency resupply $P(Y = 0)$. Based on simulation experiments, we find that Y is well-approximated by a negative binomial distribution. We estimate $P(Y = 0)$ using the approximate mean and variance of Y, assuming Y follows a negative binomial distribution. We approximate the variance of Y using

$$Var(Y) \approx \sum_{k_1=1}^{K}\sum_{k_2=1}^{K} Cov(Y_{k_1}, Y_{k_2}) = \sum_{k=1}^{K}(\Lambda_2^k - (\Lambda_1^k)^2) + \sum_{k_1<k_2\leq K}(\Lambda_2^{k_1k_2} - \Lambda_2^{k_1} - \Lambda_2^{k_2} - 2\Lambda_1^{k_1}\Lambda_1^{k_2}). \tag{3.19}$$

The mean of Y is approximated by

$$E(Y) \approx \sum_{k=1}^{K} \Lambda_1^k. \tag{3.20}$$

The experimental validation of this approach for approximating these moments and probability of zero for Z can be found in the appendix.

3.5 Methodology of Estimating the Stationary Distribution of the Number of the Emergency Ordered Units

In this section, we integrate the ideas developed in previous sections into a practical method for estimating the stationary distribution of the number of units in emergency resupply.

As suggested from simulation experiments (EC.2 in the appendix), we use a mixed distribution approach. We first define a mixed distribution with an atom at zero and a zero-truncated negative binomial distribution (i.e. a distribution with positive support only) first. We next calculate the parameters for the zero-truncated negative binomial distribution given the mean, variance and probability of zero point of the mixed distribution. We then use this mixed distribution to approximate the steady state distribution of the number of the emergency ordered units for a system involving an RSL, an ESL and a single FSL. We also extend the results to multiple FSLs with independent demand processes. Finally, we compare this approximating distribution to a distribution obtained from simulation to measure the accuracy of the proposed approximation.

3.5.1 The Reduced State Binomial Approximation Method

Let $\tilde{Z}$ denote the random variable with the distribution formed by mixing an atom at zero and a zero-truncated negative binomial distribution. The distribution has three parameters $\{p_0, \alpha, \beta\}$: $p_0 = P(\tilde{Z} = 0)$, is the probability of zero units in resupply; and $\{\alpha, \beta\}$ are the parameters of the zero-truncated negative binomial distributed $\tilde{Z}_{|\tilde{Z}>0}$. Therefore, the probability distribution function of $\tilde{Z}$

could be expressed as:

$$P(\tilde{Z} = \tilde{z}) = \begin{cases} p_0 & \text{if} \tilde{z} = 0 \\ (1 - p_0)[\frac{1}{1-\beta^\alpha}\binom{\tilde{z}+\alpha-1}{\tilde{z}}\beta^\alpha(1-\beta)^{\tilde{z}}] & \text{if} \tilde{z} > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3.21}$$

Proposition 10 *Given the mean μ, variance σ^2 and zero point probability p_0 of $\tilde{Z}$, the parameters $\{\alpha, \beta\}$ of the zero-truncated negative binomial distribution satisfy the following conditions:*

- *α is the zero point for the function $f(r)$:*

$$f(r) = \frac{1}{1 - (\frac{1+r}{\frac{\sigma^2+\mu^2}{\mu}+r})^r} r(\frac{\frac{\sigma^2+\mu^2}{\mu} - 1}{1 + r}) - \frac{\mu}{1 - p_0}, \tag{3.22}$$

that is, α satisfies $f(\alpha) = 0$,

- *β is given by:*

$$\beta = \frac{1 + \alpha}{\frac{\sigma^2+\mu^2}{\mu} + \alpha}. \tag{3.23}$$

For a system consisting of a single FSL, an RSL and an ESL, we propose the following method to approximate the limiting distribution of Z, the number of the emergency ordered units. First choose a large value for K. Then,

1. Approximate $E(Y)$ using (3.20) and $E(Z)$ using (3.18).
2. Approximate $Var(Y)$ using (3.19)and $Var(Z)$ using (3.17).
3. Compute $P(Z = 0)$ using $P(Y = 0)$, assuming Y has a negative binomial distribution; and
4. Use (3.21), the distribution of $\tilde{Z}$, to approximate the limiting distribution of Z and use Proposition 10 to compute its parameters.

We refer to this approach as the *reduced-state bimodal approximation method* (the RSB method).

For a system having H independent FSLs, let $Z^{(h)}, h = 1, 2, \ldots, H$ denote the random variable with the limiting distribution of the number of the emergency ordered units for the h^{th} FSL. Let Z^{total} denote the random variable with the limiting distribution of the total number of the emergency ordered units. Then, $Z^{total} = \sum_{h=1}^{H} Z^{(h)}$.

Since the demands arriving at the FSLs are independent, we have

$$E(Z^{total}) = \sum_{h=1}^{H} E(Z^{(h)}),$$

$$Var(Z^{total}) = \sum_{h=1}^{H} Var(Z^{(h)}),$$

and

$$P(Z^{total} = 0) = \prod_{h=1}^{H} P(Z^{(h)} = 0).$$

As with the single FSL system, we assume that the limiting distribution Z^{total} follows the mixed distribution with an atom at zero and a zero-truncated negative binomial distribution. We get an estimate of the parameters of this mixed distribution by using Proposition 10 together with the mean, variance and probability of zero point of Z^{total}, from above.

3.5.2 Simulation Comparison

In our simulation experiments, we focus on the accuracy of the approximation in estimating the right hand tail of the stationary distribution of the number of units on emergency order. This tail is the important portion of the distribution for the purpose of optimizing stock levels.

Identical Independent FSLs

First, we conduct a simulation study assuming that the demand distribution is identical for all FSLs. Let W_t denote the cumulative unit arrivals during time t for a FSL. We fix the lead time, $\tau_F = \tau_E$, equal to 7, and the arrival process mean rate $\frac{E(W_t)}{t} = \frac{\lambda}{p}$, equal to 5 per time unit at each FSL. The arrival process variance rate, $\frac{Var(W_t)}{t} = \lambda(\frac{1-p}{p^2} + (\frac{1}{p})^2) = \sigma^2$, is taken to be one of 10, 100 or 1000. The simulation runtime is $100,000$ time units when σ^2 is equal to either 10 or 100, and $1,000,000$ time units when $\sigma^2 = 1000$. The number of the FSLs is $H = 1, 3$.

To see whether including the atom at zero adds value to the quality of the approximation, we consider a simpler method which skips step 3 in the RSB method and approximates the limiting distribution of Z by a negative binomial distribution. Let us name this approximation as NB method. Figures 3.2-3.7 show the comparison between the estimated quantiles (the RSB method and the NB method) and the simulation quantiles for various combinations of the demand variance, σ^2, stock level, $S = 50, 40, 35$ (from left to right), and the number of FSLs, $H = 1, 3$. We compare our mixed distribution (straight lines with $*$ marks) with the simulated distribution (dash lines with confidence intervals $+-+$) and heuristic negative binomial distribution (dash dot lines with $\times$ marks) and use the $Q\%$-quantile z_Q as the metric, defined as

$$P(Z \leq z_Q) \geq 1 - Q.$$

We compare the quantiles for $Q = 10\%, 5\%, 2.5\%, 1\%, 0.5\%$ separately.

Overall, the estimated quantiles are quite close to the empirical results. As the variance-to-mean ratio increases, the RSB method results in better quantile estimation than the NB method. Meanwhile, the quantiles collapse together as

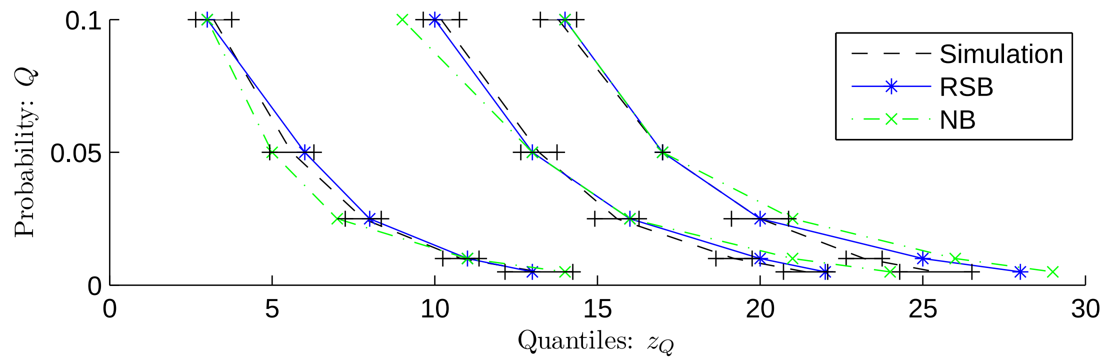

Figure 3.2: Upper Quantiles of Distribution of Units in Emergency Resupply Plots ($\sigma^2 = 10$, $H = 1$)

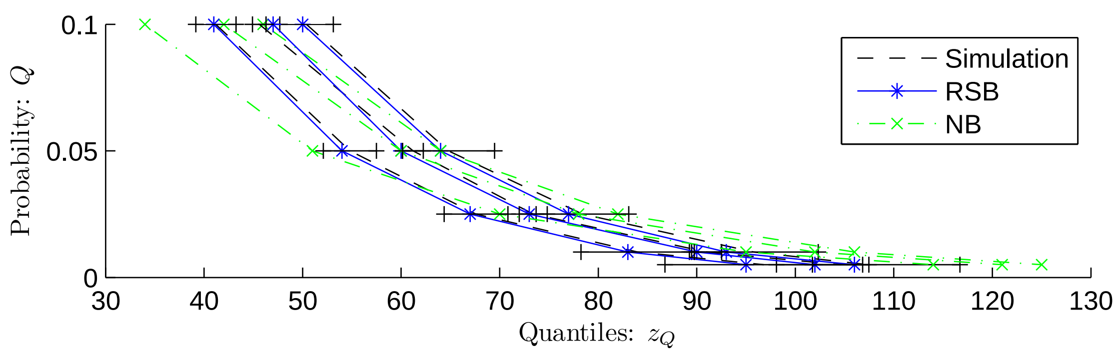

Figure 3.3: Upper Quantiles of Distribution of Units in Emergency Resupply Plots ($\sigma^2 = 100$, $H = 1$)

the variance-to-mean ratio increases, which means that the effect of the stock levels ($S = 50, 40, 35$) at the FSLs on those quantiles diminishes.

Non-Identical Independent FSLs

We conduct another simulation study assuming that there are 10 different FSLs and the means of the demand follow a Pareto curve. Let $W_t^{(l)}$ denote the cumulative unit arrivals during time t for the l^{th}-largest FSL. We fix the lead

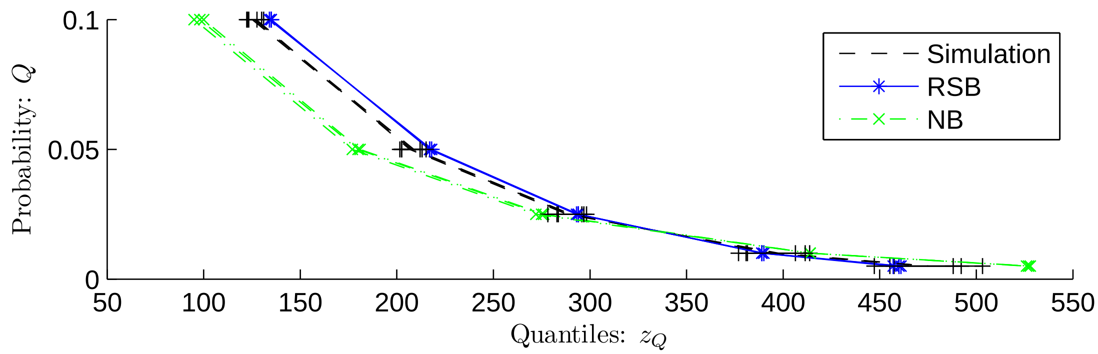

Figure 3.4: Upper Quantiles of Distribution of Units in Emergency Resupply Plots ($\sigma^2 = 1000$, $H = 1$)

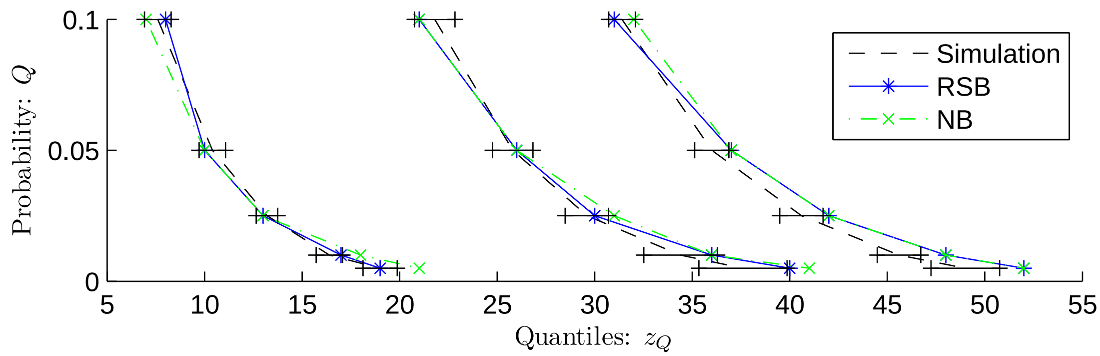

Figure 3.5: Upper Quantiles of Distribution of Units in Emergency Resupply Plots ($\sigma^2 = 10$, $H = 3$)

time to be $\tau_F = \tau_E = 7$. The total arrival process mean rate $\sum_{l=1}^{10} \frac{E(W_t^{(l)})}{t} = \sum_{l=1}^{10} \mu^{(l)} = 50$ units per time unit and the mean unit demand rates for the ten FSLs are $27.5, 12.4, 5.57, 2.51, 1.13, 0.507, 0.228, 0.103, 4.62 \times 10^{-2}, 2.08 \times 10^{-2}$, respectively. The arrival process coefficient of variation (CV) for the ten FSLs are $\{0.5, 1, 1.5, \ldots, 9.5\}$, respectively. The simulation runtime is $100,000$ time units for all cases.

We compare our mixed distribution with the simulation distribution and use

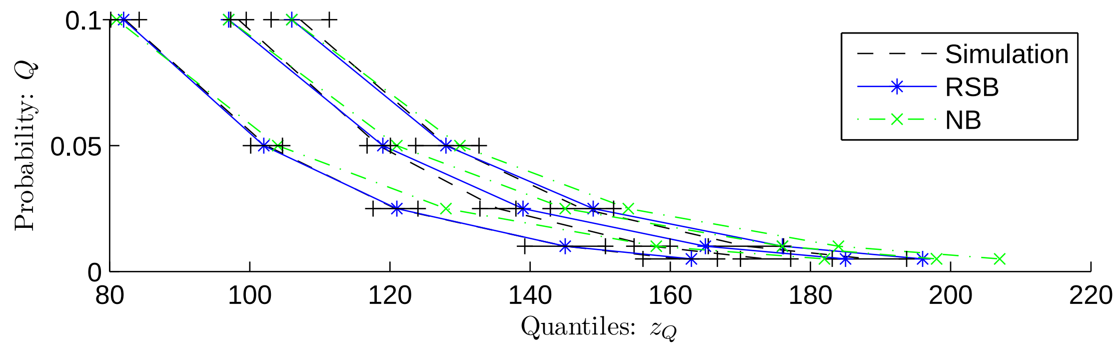

Figure 3.6: Upper Quantiles of Distribution of Units in Emergency Resupply Plots ($\sigma^2 = 100$, $H = 3$)

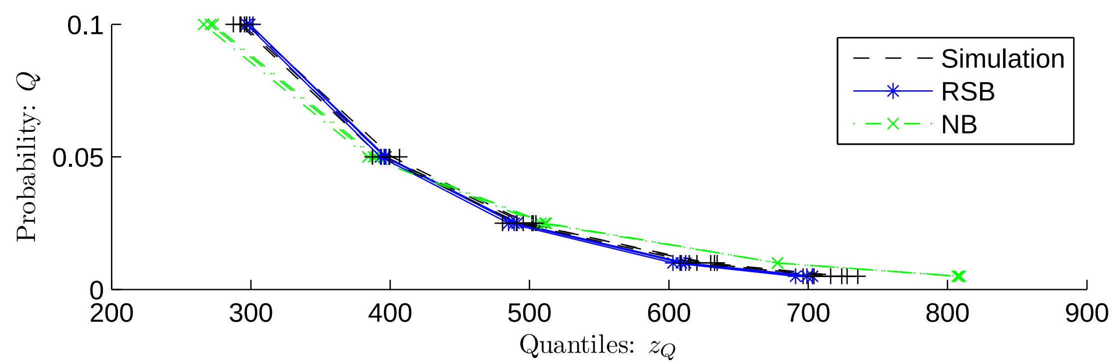

Figure 3.7: Upper Quantiles of Distribution of Units in Emergency Resupply Plots ($\sigma^2 = 1000$, $H = 3$)

the Q-quantile z_Q ($Q = 10\%, 5\%, 2.5\%, 1\%, 0.5\%$) as the metric.

Figure 3.8 shows the comparison between the estimated quantiles (the RSB method and the NB method) and the simulation quantiles for different targeted stock levels at the FSLs, $S^{(l)} = \tau\times\mu^{(l)}+\rho\sqrt{\tau}\times\sigma^{(l)}$, for $\rho = 3, 2, 1, 0$ (from left to right), $l = 1, 2, \ldots, 10$. Both the mixed distribution (RSB) and the negative binomial (NB) distribution approximations estimate the empirical quantiles quite well. The mixed distribution yields a slightly better quantile estimation as the amount

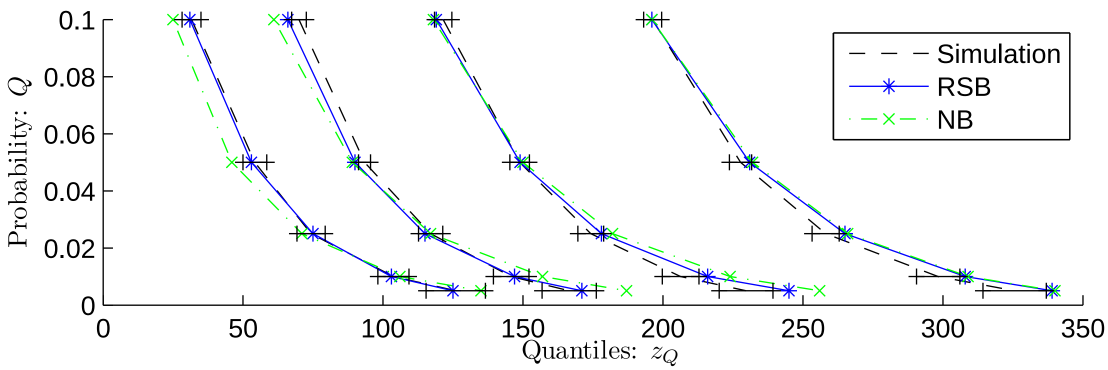

Figure 3.8: Upper Quantiles of Distribution of Units in Emergency Resupply Plots ($S^{(l)} = \tau \times \mu^{(l)} + \rho\sqrt{\tau} \times \sigma^{(l)}$for $l = 1, 2, \ldots, 10$)

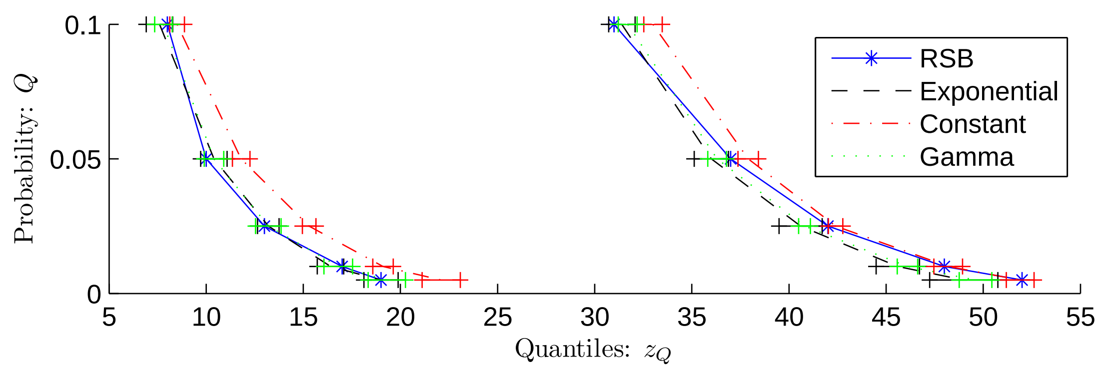

Figure 3.9: Comparison of the Upper Quantiles of Distribution of Units in Emergency Resupply with Different Lead Time Distributions ($\sigma^2 = 10$, $H = 3$)

of safety stock.

Non-Exponential Lead Time Identical Independent FSLs

Chen.*et al.* (2010) prove that the distribution of the number of units on order at the FSLs (regular replenishment) is insensitive to the form of the lead time distribution as long as the lead times are independent and identically distributed

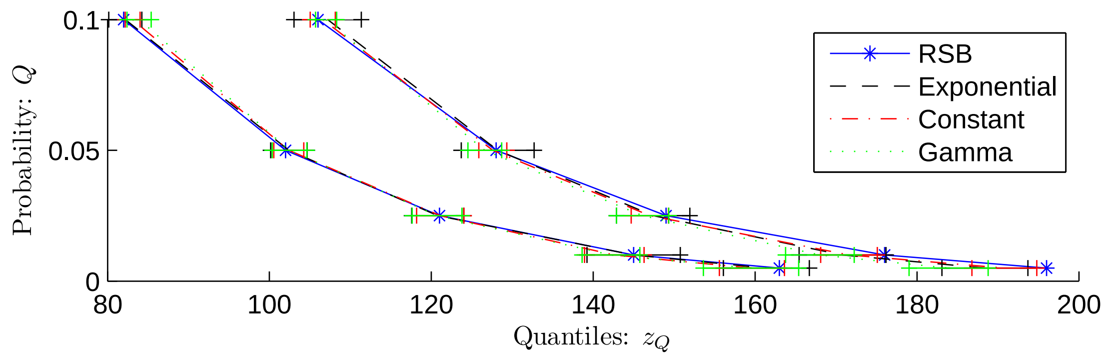

Figure 3.10: Comparison of the Upper Quantiles of Distribution of Units in Emergency Resupply with Different Lead Time Distributions ($\sigma^2 = 100$, $H = 3$)

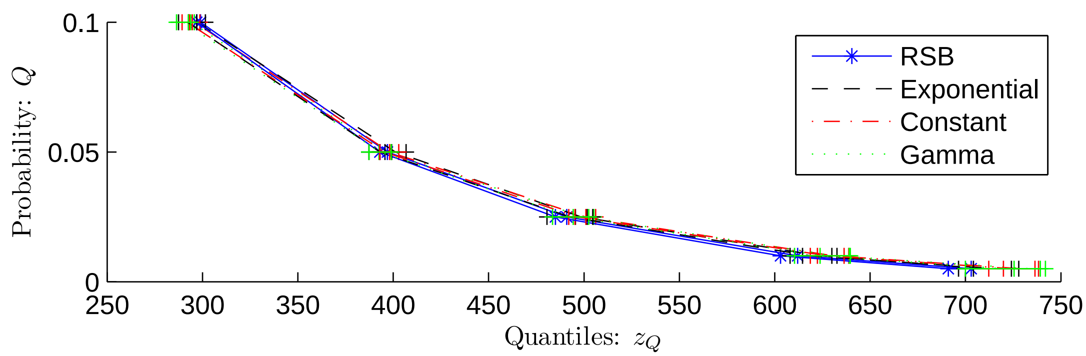

Figure 3.11: Comparison of the Upper Quantiles of Distribution of Units in Emergency Resupply with Different Lead Time Distributions ($\sigma^2 = 1000$, $H = 3$)

with the same mean. However, as evident from our simulations experiments, the analogous property does not hold for the number of emergency ordered units at the ESL. To check the sensitivity of the results to the choice of the lead time distribution, we conduct experiments with three different lead time distributions. Here the distributions of the lead time are taken to be, alternatively, the exponential distribution ("−−−"), the constant ("−·−"), and the gamma distribution ("···") with VTMR equal to 5. Figures 3.9-3.11 show the comparison of the

RSB (straight lines) approach estimates with the empirical results. The demand variance σ^2 is equal to $10, 100$ or 1000 and the stock levels S equal either 50 or 35 (from left to right) , and the number of FSLs H is equal to 3. We observed that as the distribution of the lead time changes, the quantiles change but the changes are small. As the variance-to-mean ratio of the demand increases, our RSB estimates fit quite well.

Consequently, although our method is developed for exponentially-distributed lead times, it appears to work well for any finite mean lead time distribution, provided lead times are independent and identically distributed. Alfredsson and Verrijdt (1999) have a similar finding in a two-echelon model with lateral transshipment, namely, that performance measures are relatively insensitive to the form of the lead time distribution.

3.6 Conclusion

For the special case of stuttering Poisson demand, exponentially distributed lead times, and $(S-1, S)$ replenishment policies we have investigated the stationary distribution of the number of units in emergency resupply for a system composed of an RSL, an ESL and multiple FSLs. We use the lost sales model as the basis for representing such emergency order systems. We develop exact expressions for the first and second moments of the number of outstanding emergency orders. Using an approximating Markov chain, we provide a tractable method for calculating these moments, and we use them to estimate the mean, variance, and probability of the zero point of the number of emergency ordered units at the ESL. Our approach, called the reduced state bimodal approximation method (RSB), uses these estimated statistics to approximate the distribution of the number of units in emergency resupply as a zero-truncated negative binomial distribution mixed with an atom at zero. Quantile analysis using simulation demonstrates that these approximations are quite accurate.

In a companion paper, we develop an optimization algorithm for setting stock levels in a system with both field service locations and an emergency stocking location.

Acknowledgements

The authors are indebted to Gennady Samorodnitsky for contributions to the approach taken in sections 3 and 4.3.

3.A Experimental Validation

When calculating the statistics of the emergency order distribution, we assume that the conditional stationary distribution of the number of outstanding orders at the ESL depends only on the on-hand inventory quantities at the FSL. When estimating $P(Z = 0)$, we further assume the number of emergency orders follows a negative binomial distribution. A simulation study is employed to test the quality of this approximation.

Let W_t denote the cumulative unit arrivals through time t. For these simulations, we fix the expected lead times, $\tau_F = \tau_E$, equal to 7, and the arrival process mean rate, $\frac{E(W_t)}{t} = \frac{\lambda}{p}$, equal to 5 per time unit. The arrival process variance rate

Table 3.1: Comparison of Mean, Variance and Zero Point Estimates with Simulation for $VTMR = \frac{10}{5} = 2$:

	Estimation	Simulation		Estimation	Simulation		Estimation	Simulation	
S	$E(Z)$	$\hat{\mu}_Z$	95% CI for μ_Z	$Var(Z)$	$\hat{\sigma}^2_Z$	95% CI for σ^2_Z	$P(Z=0)$	$\hat{P}(Z=0)$	95% CI for $P(Z=0)$
1	34.06	33.93	[33.74 34.11]	69.84	68.32	[66.54 70.11]	0	0	[0 0]
5	30.33	30.35	[30.13 30.58]	68.97	69.70	[68.22 71.18]	0	0	[0 0]
10	25.73	25.72	[25.52 25.92]	67.16	67.25	[65.65 68.84]	0	0	[0 0]
20	16.96	16.81	[16.58 17.05]	59.57	59.32	[57.92 60.72]	0	0	[0 0]
30	9.25	9.31	[9.08 9.54]	43.51	44.62	[42.63 46.61]	0.05	0.065	[0.060 0.070]
40	3.61	3.61	[3.49 3.74]	20.83	20.57	[19.59 21.54]	0.36	0.38	[0.37 0.39]
50	0.83	0.85	[0.82 0.87]	4.94	4.99	[4.73 5.24]	0.80	0.80	[0.79 0.81]

$\frac{Var(W_t)}{t} = \lambda(\frac{1-p}{p^2}+(\frac{1}{p})^2) = \sigma^2$ is taken to be either 10, 100 or 1000. The duration of the simulation experiment is 100,000 time units when $\sigma^2 = 10$ or 100 and 1,000,000 time units when $\sigma^2 = 1000$.

For the summation limit K used to estimate $E(Z) \approx \sum_{k=1}^{K} kY_k$ and

$$Var(Z) \approx \sum_{k_1=1}^{K}\sum_{k_2=1}^{K} k_1k_2Cov(Y_{k_1}, Y_{k_2}),$$

we let $K \equiv \frac{Var(X_t)}{t} = \sigma^2$ (but no less than 100). With this value of K, the probability of an order overshoot is given by

$$P(X > K) = \bar{P}_K = (1-p)^K = (1 - \frac{2}{VTMR+1})^K = (1 - \frac{2}{\frac{\sigma^2}{\lambda p}+1})^{\sigma^2} \sim 2\%,$$

for all chosen values of σ^2.

Let $\hat{\mu}_Z$, $\hat{\sigma}^2_Z$, $\hat{P}(Z=0)$ represent the simulation sample mean, sample variance, and sample zero point probability, respectively. The corresponding confidence interval (CI) for each statistic is calculated by batch means. Tables 1-3 show the comparisons of the estimates (Section 5) and simulation results when the $VTMR$ varies from 2 ($\sigma^2 = 10$) to 200 ($\sigma^2 = 1000$). The estimates in most cases are very close to the sample means, variances and zero point probabilities; but, the

Table 3.2: Comparison of Mean ,Variance and Zero Point Estimates with Simulation for $VTMR = \frac{100}{5} = 20$:

	Estimation	Simulation		Estimation	Simulation		Estimation	Simulation	
S	$E(Z)$	$\hat{\mu}_Z$	95% CI for μ_Z	$Var(Z)$	$\hat{\sigma}_Z^2$	95% CI for σ_Z^2	$P(Z=0)$	$\hat{P}(Z=0)$	95% CI for $P(Z=0)$
1	34.74	34.54	[33.96 35.12]	697.92	702.28	[674.19 730.37]	0.05	0.04	[0.04 0.05]
5	33.00	33.05	[32.64 33.45]	692.72	693.10	[672.63 713.58]	0.08	0.08	[0.08 0.09]
10	30.36	30.75	[30.29 31.22]	678.12	679.53	[651.25 707.80]	0.14	0.13	[0.13 0.14]
20	24.99	25.03	[24.37 25.70]	629.36	630.80	[606.63 654.97]	0.25	0.25	[0.25 0.26]
30	20	20.10	[19.33 20.87]	561.41	559.25	[534.77 583.73]	0.38	0.37	[0.36 0.39]
40	15.61	15.77	[15.33 16.20]	481.43	495.65	[467.94 523.35]	0.50	0.50	[0.49 0.51]
50	11.86	11.96	[11.47 12.45]	396.74	406.22	[380.98 431.47]	0.62	0.62	[0.61 0.63]

Table 3.3: Comparison of Mean ,Variance and Zero Point Estimates with Simulation for $VTMR = \frac{1000}{5} = 200$:

	Estimation	Simulation		Estimation	Simulation		Estimation	Simulation	
S	$E(Z)$	$\hat{\mu}_Z$	95% CI for μ_Z	$Var(Z)$	$\hat{\sigma}_Z^2$	95% CI for σ_Z^2	$P(Z=0)$	$\hat{P}(Z=0)$	95% CI for $P(Z=0)$
1	34.98	35.24	[34.22 36.25]	6980.69	7103.81	[6823.06 7384.56]	0.71	0.71	[0.70 0.71]
5	34.93	34.43	[33.62 35.24]	6980.51	6938.71	[6580.98 7296.45]	0.72	0.72	[0.72 0.72]
10	34.81	34.74	[34.06 35.41]	6979.48	6961.75	[6801.59 7121.90]	0.73	0.73	[0.72 0.73]
20	34.37	33.70	[32.95 34.45]	6972.47	6754.92	[6532.43 6977.40]	0.75	0.75	[0.75 0.76]
30	33.72	33.87	[33.21 34.53]	6955.99	7070.94	[6862.19 7279.70]	0.77	0.77	[0.77 0.77]
40	32.93	33.23	[32.47 33.98]	6927.80	6960.01	[6676.72 7243.31]	0.79	0.79	[0.78 0.79]
50	32.02	32.03	[31.25 32.80]	6886.67	6840.09	[6604.32 7075.87]	0.80	0.80	[0.80 0.81]

confidence intervals are large in some cases, particulary for $Var(Z)$ as the VTMR increases.

We conclude that the proposed method for approximating the mean, variance and the probability of zero units in emergency resupply should perform well under the assumption of a stuttering Poisson demand process and exponentially-distributed lead times.

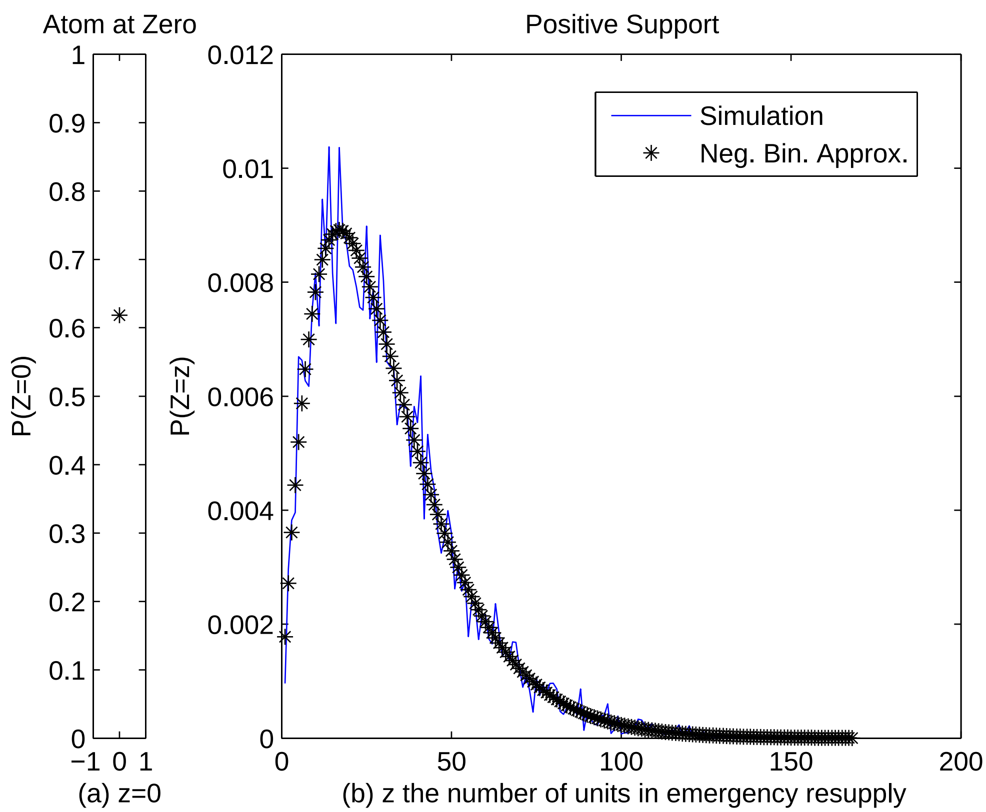

Figure 3.12: (a) Atom at Zero (b) Simulation VS Zero-truncated Negative Binomial Approximation on Positive Support (VTMR=20)

3.B Simulation of Shape of the Stationary Distribution of the Number of the Emergency Ordered Units

We use simulation experiments to explore the shape of the stationary distribution of the number of units on emergency order. From these simulation results with different parameters, we observe that the stationary distribution has an atom at zero and the distribution on positive support appears to be a truncated-at-zero negative binomial distribution.

For example, let W_t denote the cumulative unit arrivals during time t for a

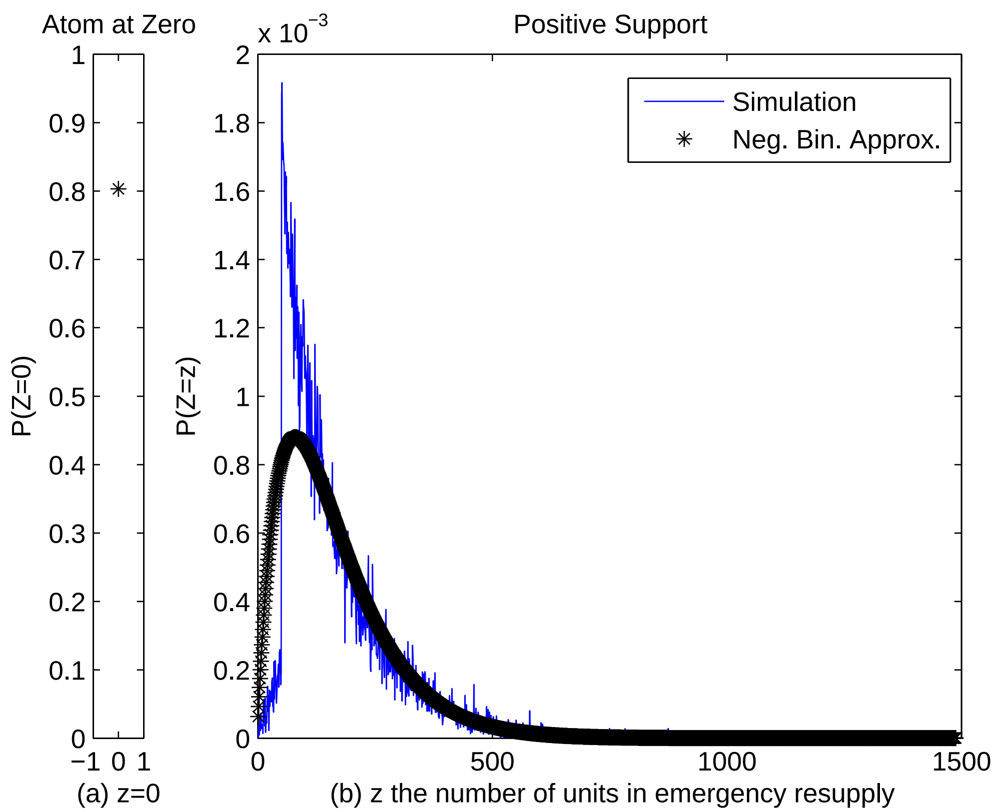

Figure 3.13: (a) Atom at Zero (b) Simulation VS Zero-truncated Negative Binomial Approximation on Positive Support (VTMR=200)

FSL. Let us look at a system with one FSL, an ESL, and an RSL. The lead time, τ, is equal to 7, and the arrival process mean rate $\frac{E(W_t)}{t} = \frac{\lambda}{p}$, is equal to 5 per time unit at the FSL. The arrival process variance rate, $\frac{Var(W_t)}{t} = \lambda(\frac{1-p}{p^2} + (\frac{1}{p})^2) = \sigma^2$, is taken to be 100. The order-up-to level of the FSL is $S = 50$. The simulation runtime is $100,000$ time units. From the simulation results, there is clearly an atom at zero that averages to $\hat{P}(Z = 0) = 0.62$ (Figure 3.12(a)). Conditioned on this average, as well as using the mean and variance of the simulation results, we estimate the parameters of a negative binomial distribution. Figure 3.12(b) displays both the empirical distribution on positive support, from the simulation, together with this fitted distribution. As the variance to mean ratio of the

Table 3.4: Upper Quantiles of Distribution of Units in Emergency Resupply ($\sigma^2 = 10, H = 1$):

	$S = 35$				$S = 40$				$S = 50$			
Q	RSB	NB	Simu.	95%CI	RSB	NB	Simu.	95%CI	RSB	NB	Simu.	95%CI
10%	14	14	13.8	[13.2 14.3]	10	9	10.2	[9.6 10.8]	3	3	3.2	[2.6 3.8]
5%	17	17	17	[17 17]	13	13	13.2	[12.6 13.7]	6	5	5.6	[4.9 6.3]
2.5%	20	21	20	[19.1 20.9]	16	16	15.6	[14.9 16.3]	8	7	7.8	[7.2 8.4]
1%	25	26	23.2	[22.6 23.8]	20	21	19.2	[18.6 19.8]	11	11	10.8	[10.2 11.4]
0.5%	28	29	25.4	[24.3 26.5]	22	24	21.4	[20.7 22.1]	13	14	13.2	[12.1 14.2]

Table 3.5: Upper Quantiles of Distribution of Units in Emergency Resupply ($\sigma^2 = 100, H = 1$):

	$S = 35$				$S = 40$				$S = 50$			
Q	RSB	NB	Simu.	95%CI	RSB	NB	Simu.	95%CI	RSB	NB	Simu.	95%CI
10%	50	46	50.4	[47.7 53.1]	47	42	45.6	[44.9 46.3]	41	34	41.2	[39.2 43.2]
5%	64	64	64.8	[60.1 69.5]	60	60	61.2	[60.1 62.2]	54	51	554.8	[52.1 57.5]
2.5%	77	82	78.4	[73.7 83.1]	73	78	73.4	[72.0 74.8]	67	70	67.6	[64.4 70.8]
1%	93	106	95.8	[89.3 102.3]	90	102	91	[89.5 92.5]	83	85	84	[78.2 89.8]
0.50%	106	125	107.4	[98.1 116.7]	102	121	104.6	[101.7 107.5]	95	114	96.8	[86.8 114]

demand increases, the truncated-at-zero negative binomial distribution approximation is not as accurate. However, it still approximates well to the upper tail quantiles, which is the proportion of the distribution that affects the to optimal target stock level at the ESL. Figure 3.13 is such an example with the same parameters as the previous case except the arrival process variance rate σ^2 is 1000.

Table 3.6: Upper Quantiles of Distribution of Units in Emergency Resupply (σ^2 = 1000, H = 1):

	S = 35				S = 40				S = 50			
Q	RSB	NB	Simu.	95%CI	RSB	NB	Simu.	95%CI	RSB	NB	Simu.	95%CI
10%	135	100	125.2	[122.8 127.6]	135	98	127.2	[123.3 131.1]	134	95	126.2	[122.3 130.1]
5%	217	181	207.8	[202.5 213.1]	217	180	206.8	[201.6 212.0]	218	177	208.8	[202.4 215.2]
2.5%	293	276	289.4	[283.0 295.8]	294	275	288.2	[278.2 298.2]	294	272	290.2	[283.6 296.8]
1%	390	414	393.6	[380.8 406.4]	389	414	397.6	[381.5 413.7]	389	414	394.2	[377.0 411.4]
0.5%	461	526	471.6	[455.2 488.0]	460	527	474.8	[457.4 492.2]	457	528	475.2	[447.2 503.2]

Table 3.7: Upper Quantiles of Distribution of Units in Emergency Resupply (σ^2 = 10, H = 3):

	S = 35				S = 40				S = 50			
Q	RSB	NB	Simu.	95%CI	RSB	NB	Simu.n	95%CI	RSB	NB	Simu.	95%CI
10 %	31	32	31.4	[30.7 32.1]	21	21	21.8	[20.8 22.8]	8	7	7.6	[6.9 8.3]
5 %	37	37	36.0	[35.1 36.9]	26	26	25.8	[24.8 26.8]	10	10	10.4	[9.7 11.1]
2.5 %	42	42	40.6	[39.5 41.7]	30	31	29.6	[28.5 30.7]	13	13	13.2	[12.6 13.8]
1 %	48	48	45.6	[44.5 46.7]	36	36	34.4	[32.5 36.3]	17	18	16.4	[15.7 17.1]
0.5 %	52	52	49.0	[47.2 50.8]	40	41	37.6	[35.3 39.9]	19	21	19.0	[18.1 19.9]

Table 3.8: Upper Quantiles of Distribution of Units in Emergency Resupply (σ^2 = 100, H = 3):

	S = 35				S = 40				S = 50			
Q	RSB	NB	Simu.	95%CI	RSB	NB	Simu.	95%CI	RSB	NB	Simu.	95%CI
10 %	106	106	107.2	[103.0 111.4]	97	97	98.4	[97.3 99.5]	82	81	82.2	[80.2 84.2]
5 %	128	130	128.2	[123.7 132.7]	119	121	118.4	[116.7 120.1]	102	104	102.4	[100.1 104.7]
2.5 %	149	154	147.4	[142.9 151.9]	139	145	135.4	[132.8 138.0]	121	128	120.8	[117.6 124.0]
1 %	176	184	170.8	[165.4 176.2]	165	176	157.4	[154.8 160.0]	145	158	145.0	[139.2 150.8]
0.5 %	196	207	188.4	[183.1 193.7]	185	198	173.6	[170.0 177.2]	163	182	161.4	[156.1 166.7]

Table 3.9: Upper Quantiles of Distribution of Units in Emergency Resupply ($\sigma^2 = 1000$, $H = 3$):

	$S = 35$				$S = 40$				$S = 50$			
Q	RSB	NB	Simu.	95%CI	RSB	NB	Simu.	95%CI	RSB	NB	Simu.	95%CI
10 %	299	273	297.2	[292.9 301.5]	299	271	293.8	[292.2 295.4]	297	266	292.2	[287.4 297.0]
5 %	397	390	401.2	[395.7 406.7]	396	388	396.6	[393.6 399.6]	393	384	393.2	[387.3 399.1]
2.5 %	491	512	498.6	[495.5 501.7]	489	511	496.6	[490.6 502.6]	485	508	492.6	[480.6 504.6]
1 %	612	678	623.6	[614.5 632.7]	609	678	627.4	[620.2 634.6]	603	678	619.0	[608.0 630.0]
0.5 %	703	807	710.2	[696.3 724.1]	699	808	725.8	[716.1 735.5]	691	809	716.0	[704.1 727.9]

3.C Simulation Numerical Results

3.C.1 Identical Independent FSLs

Tables 3.4-3.9 show the comparison between the estimated quantiles, the simulation quantiles and the negative binomial distribution approximate quantiles for various combinations of demand variance, σ^2, stock level, S, and the number of independent FSLs, H, which have identical demand distributions.

In half of the cases, the estimated quantile lie within the 95% confidence interval from the simulation and in the other cases, the estimated quantiles are close to them. The results also show that when we have multiple FSLs with independent and identical demand processes, the difference between the estimated quantile and the simulation sample mean does not increase as the number of FSLs increases.

Table 3.10: Upper Quantiles of Distribution of Units in Emergency Resupply ($S^{(l)} = \tau \times \mu^{(l)} + \rho\sqrt{\tau} \times \sigma^{(l)}$for $l = 1, 2, \ldots, 10$):

	$\rho = 0$				$\rho = 1$			
Q	RSB	NB	Simulation	95%CI	RSB	NB	Simulation	95%CI
10 %	196	196	196.3	[193.1 199.5]	119	118	121.5	[118.4 124.6]
5 %	231	232	227.7	[223.7 231.7]	149	150	148.8	[145.3 152.3]
2.5 %	265	266	258.1	[253.3 262.9]	178	182	174.2	[169.6 178.8]
1 %	308	309	298.3	[290.6 306.0]	216	224	206.3	[199.7 212.9]
0.5 %	339	340	325.6	[314.3 336.9]	245	256	229.7	[220.1 239.3]

Table 3.11: Upper Quantiles of Distribution of Units in Emergency Resupply ($S^{(l)} = \tau \times \mu^{(l)} + \rho\sqrt{\tau} \times \sigma^{(l)}$ for $l = 1, 2, \ldots, 10$):

	$\rho = 0$				$\rho = 1$			
Q	RSB	NB	Simulation	95%CI	RSB	NB	Simulation	95%CI
10 %	66	61	70	[67.4 72.6]	31	25	31.6	[28.2 35.0]
5 %	90	89	93.4	[91.3 95.5]	53	46	54.2	[50.0 58.4]
2.5 %	115	117	117	[112.7 121.3]	75	71	74.3	[69.3 79.3]
1 %	147	157	145.8	[139.5 152.1]	103	106	103.7	[98.1 109.3]
0.5 %	171	187	166.5	[156.8 176.2]	125	135	126	[115.4 136.6]

Non-identical Independent FSLs

Table EC.10-EC.11 show the comparison between the estimated quantiles, the simulation quantiles and the negative binomial distribution approximate quantiles for different targeted stock level stock, $S^{(l)}$, for $l = 1, 2, \ldots, 10$. In this case, the demand processes are not identical among the FSLs.

The estimated quantile either lies within or is very close to the 95% confidence interval from the simulation. Therefore, we believe that our approximation is useful and effective, especially for the upper tails of the distribution.

Table 3.12: Comparison of the Upper Quantiles of Distribution of Units in Emergency Resupply with Different Lead Time Distributions ($\sigma^2 = 10$, $H = 3$):

$S = 35$		Exponential		Constant		Gamma	
Q	Estimated	Simulation	95%CI	Simulation	95%CI	Simulation	95%CI
10 %	31	31.4	[30.7 32.1]	33	[32.5 33.5]	31.7	[31.2 32.2]
5 %	37	36	[35.1 36.9]	37.9	[37.4 38.4]	36.3	[35.8 36.8]
2.5 %	42	40.6	[39.5 41.7]	42.4	[42.0 42.8]	40.8	[40.5 41.1]
1 %	48	45.6	[44.5 46.7]	48.2	[47.5 48.9]	46.1	[45.6 46.6]
0.5 %	52	49	[47.2 50.8]	51.9	[51.2 52.6]	49.6	[48.8 50.4]

Table 3.13: Comparison of the Upper Quantiles of Distribution of Units in Emergency Resupply with Different Lead Time Distributions ($\sigma^2 = 10$, $H = 3$):

$S = 50$		Exponential		Constant		Gamma	
Q	Estimated	Simulation	95%CI	Simulation	95%CI	Simulation	95%CI
10 %	8	7.6	[6.9 8.3]	8.5	[8.1 8.9]	7.8	[7.3 8.3]
5 %	10	10.4	[9.7 11.1]	11.8	[11.3 12.3]	10.4	[9.9 10.9]
2.5 %	13	13.2	[12.6 13.8]	15.3	[15.0 15.6]	13.2	[12.5 13.9]
1 %	17	16.4	[15.7 17.1]	19.1	[18.6 19.6]	16.8	[16.1 17.5]
0.5 %	19	19	[18.1 19.9]	22.1	[21.1 23.1]	19.3	[18.3 20.3]

Table 3.14: Comparison of the Upper Quantiles of Distribution of Units in Emergency Resupply with Different Lead Time Distributions ($\sigma^2 = 100$, $H = 3$):

$S = 35$		Exponential		Constant		Gamma	
Q	Estimated	Simulation	95%CI	Simulation	95%CI	Simulation	95%CI
10 %	106	107.2	[103.0 111.4]	106.6	[105.0 108.2]	107	[105.7 108.3]
5 %	128	128.2	[123.7 132.7]	127.6	[125.9 129.3]	126.6	[124.5 128.7]
2.5 %	149	147.4	[142.9 151.9]	147	[144.7 149.3]	146.1	[142.9 149.3]
1 %	176	170.8	[165.4 176.2]	171.6	[168.1 175.1]	168	[163.8 172.2]
0.5 %	196	188.4	[183.1 193.7]	190.8	[186.8 194.8]	183.9	[179.0 188.8]

Table 3.15: Comparison of the Upper Quantiles of Distribution of Units in Emergency Resupply with Different Lead Time Distributions ($\sigma^2 = 100$, $H = 3$):

$S = 50$		Exponential		Constant		Gamma	
Q	Estimated	Simulation	95%CI	Simulation	95%CI	Simulation	95%CI
10 %	82	82.2	[80.2 84.2]	83.1	[82.2 84.0]	83.9	[82.4 85.4]
5 %	102	102.4	[100.1 104.7]	102.4	[100.5 104.3]	102.5	[100.4 104.6]
2.5 %	121	120.8	[117.6 124.0]	121.1	[118.2 124.0]	120.7	[117.6 123.8]
1 %	145	145	[139.2 150.8]	142.6	[138.9 146.3]	142.2	[138.6 145.8]
0.5 %	163	161.4	[156.1 166.7]	159.6	[155.6 163.6]	159.5	[153.6 165.4]

Table 3.16: Comparison of the Upper Quantiles of Distribution of Units in Emergency Resupply with Different Lead Time Distributions ($\sigma^2 = 1000$, $H = 3$):

$S = 35$		Exponential		Constant		Gamma	
Q	Estimated	Simulation	95%CI	Simulation	95%CI	Simulation	95%CI
10 %	299	297.2	[292.9 301.5]	295.9	[293.1 298.7]	294.2	[292.6 295.8]
5 %	397	401.2	[395.7 406.7]	398	[393.1 402.9]	395.8	[392.6 399.0]
2.5 %	491	498.6	[495.5 501.7]	500.7	[495.2 506.2]	499.8	[494.4 505.2]
1 %	612	623.6	[614.5 632.7]	630.6	[622.4 638.8]	631.4	[623.8 639.0]
0.5 %	703	710.2	[696.3 724.1]	729.1	[719.9 738.3]	733.7	[725.5 741.9]

Table 3.17: Comparison of the Upper Quantiles of Distribution of Units in Emergency Resupply with Different Lead Time Distributions ($\sigma^2 = 1000$, $H = 3$):

$S = 50$		Exponential		Constant		Gamma	
Q	Estimated	Simulation	95%CI	Simulation	95%CI	Simulation	95%CI
10 %	297	292.2	[287.4 297.0]	292	[289.3 294.7]	290.1	[286.4 293.8]
5 %	393	393.2	[387.3 399.1]	395.5	[392.9 398.1]	393.1	[387.3 398.9]
2.5 %	485	492.6	[480.6 504.6]	496.5	[492.0 501.0]	493	[483.5 502.5]
1 %	603	619	[608.0 630.0]	627.5	[618.7 636.3]	625	[610.4 639.6]
0.5 %	691	716	[704.1 727.9]	724.3	[712.2 736.4]	719.6	[700.0 739.2]

Non-exponential Lead Time Identical Independent FSLs

Table EC.12-EC.17 show the comparison of the RSB approach estimates with the empirical results.

3.D Proofs

Proof of Proposition 6:

We first show the main convergence result. To prove this proposition, it suffices to show that

$$\limsup_{\substack{l\to\infty \\ h\geq 0}} E(\sum_{k=l}^{l+h} kY_k)^2 = 0.$$

Define $\bar{Y}_k$ as the random variable for the steady state number of outstanding emergency orders of size k when the target stock level S at FSL is 0. This means that any order that arrives at the FSL is immediately ordered from the ESL. It is easily seen that $\bar{Y}_k$ is stochastically larger than Y_k for all k. Thus, for all $S \geq 0$,

$$E(\sum_{k=l}^{l+h} kY_k)^2 \leq E(\sum_{k=l}^{l+h} k\bar{Y}_k)^2,$$

fir all combinations of l and h.

For $S = 0$, the system could be partitioned into an infinite number of independent sub-systems based on order size as shown in Resnick (2005 p.321). In the k^{th} sub-system, where the order size is always k, the arrivals are Poisson distributed with rate $\lambda p(1-p)^{k-1}$. So the steady state $\bar{Y}_k$ is Poisson distributed with rate $\frac{\lambda p(1-p)^{k-1}}{\mu}$ and the $\{\bar{Y}_k,\ k = 1, 2, \ldots\}$ are independent.

Therefore, for all $h > 0$,

$$\begin{aligned}
\sup E(\textstyle\sum_{k=l}^{l+h} kY_k)^2 &\leq E(\textstyle\sum_{k=l}^{l+h} k\bar{Y}_k)^2 \\
&= Var(\textstyle\sum_{k=l}^{l+h} k\bar{Y}_k) + (E\textstyle\sum_{k=l}^{l+h} k\bar{Y}_k)^2 \\
&= \textstyle\sum_{k=l}^{l+h} k^2 \frac{\lambda p(1-p)^{k-1}}{\mu} + (\textstyle\sum_{k=l}^{l+h} k\frac{\lambda p(1-p)^{k-1}}{\mu})^2 \\
&\to 0 \text{ as } l \to \infty.
\end{aligned}$$

Since the first, second and third moments for the Poisson distribution are all finite, for any k and $z = 1, 2, 3$,

$$\Lambda_z^k \leq E[\bar{Y}_k^z] < \infty.$$

Similarly, since $(\sum_{k=l}^{l+h} Y_k)^2 \leq (\sum_{k=l}^{l+h} kY_k)^2$, we have

$$\limsup_{\substack{l\to\infty \\ h\geq 0}} E(\sum_{k=l}^{l+h} Y_k)^2 \leq \limsup_{\substack{l\to\infty \\ h\geq 0}} E(\sum_{k=l}^{l+h} kY_k)^2 = 0.$$

■

Proof of Proposition 7:

Let $\tilde{A}$ denote the infinitesimal generator of the continuous time Markov chain for the k^{th} system. Then, for $(i, y) \in V \times N^+$,

$$\begin{aligned}
\tilde{A}_{(i,y),(j,y)} &= A_{i,j}, \text{for } j \neq i; \\
\tilde{A}_{(i,y),(i,y+1)} &= \rho_k(i); \\
\tilde{A}_{(i,y),(i,y-1)} &= y\mu; \\
\tilde{A}_{(i,y),(i,y)} &= -(y\mu + \rho_k(i) - A_{i,i});
\end{aligned} \tag{3.24}$$

and all other elements of $\tilde{A}$ are 0. Define $\psi_{i,-1} = 0$ for all $i \in V$. For any $(i, y) \in V \times N^{+}$, by employing the balance equations corresponding to the Markov chain and substituting for (3.24), we get

$$\begin{aligned} 0 &= \textstyle\sum_{j \neq i} \psi_{j,y} \tilde{A}_{(j,y),(i,y)} + \psi_{i,y+1} \tilde{A}_{(i,y+1),(i,y)} + \psi_{i,y-1} \tilde{A}_{(i,y-1),(i,y)} + \psi_{i,y} \tilde{A}_{(i,y),(i,y)} \\ &= \textstyle\sum_{j \neq i} \psi_{j,y} A_{j,i} + \psi_{i,y+1}(y+1)\mu + \psi_{i,y-1}\rho_k(i) - \psi_{i,y}(y\mu + \rho_k(i) - A_{i,i}). \end{aligned} \tag{3.25}$$

Since we assume Λ_1^k and Λ_2^k are finite, multiplying (3.25) by y and summing over y we get

$$\begin{aligned} 0 &= \textstyle\sum_{y=0}^{\infty} \sum_{j \neq i} \psi_{j,y} y A_{j,i} + \sum_{y=0}^{\infty} \psi_{i,y+1}(y+1)y\mu + \sum_{y=0}^{\infty} \psi_{i,y-1} y \rho_k(i) \\ &\quad - \textstyle\sum_{y=0}^{\infty} \psi_{i,y}(y^2\mu + y\rho_k(i) - yA_{i,i}) \\ &= \textstyle\sum_{j \in V} A_{j,i} \sum_{y=0}^{\infty} \psi_{j,y} y + \sum_{y=0}^{\infty} \psi_{i,y+1}(y+1)^2\mu - \sum_{y=0}^{\infty} \psi_{i,y+1}(y+1)\mu \\ &\quad + \textstyle\sum_{y=0}^{\infty} \psi_{i,y-1} y \rho_k(i) - \sum_{y=0}^{\infty} \psi_{i,y} y \rho_k(i) - \sum_{y=0}^{\infty} \psi_{i,y} y^2 \mu \\ &= \textstyle\sum_{j \in V} A_{j,i} \Lambda_{1,j}^k - \Lambda_{1,i}^k \mu + \sum_{y=0}^{\infty} \psi_{i,y-1} \rho_k(i) \\ &= \textstyle\sum_{j \in V} A_{j,i} \Lambda_{1,j}^k - \Lambda_{1,i}^k \mu + \xi_i \rho_k(i), \end{aligned}$$

as claimed in (3.5).

In matrix form

$$(\Lambda_{1,i}^k)_{1 \times |V|} \cdot (A - \mu I_{|V| \times |V|}) = -(\xi_i \rho_k(i))_{1 \times |V|}.$$

We claim $A - \mu I_{|V| \times |V|}$ is always non-singular. The proof found in Resnick(2005 p.407) can be modified slightly to obtain the desired result. This leads to (3.6).

In addition, since $\sum_{i\in V} A_{j,i} = 0$ and V is a finite state space, summing (3.5) for $i \in V$ yields

$$0 = \sum_{i\in V}\sum_{j\in V} A_{j,i}\Lambda^k_{1,j} - \sum_{i\in V}\Lambda^k_{1,i}\mu + \sum_{i\in V}\xi_i\rho_k(i) = -\Lambda^k_1\mu + \sum_{i\in V}\xi_i\rho_k(i).$$

which leads to (3.7).

For Λ^k_2, the idea is similar. Since Λ^k_2, Λ^k_3 are finite, $\Lambda^k_{2,i}$ and $\Lambda^k_{3,i}$ are finite for any i. Multiplying (3.25) by y^2 and summing over y, we get:

$$\begin{aligned}
0 &= \textstyle\sum_{y=0}^{\infty}\sum_{j\neq i}\psi_{j,y}y^2A_{j,i} + \sum_{y=0}^{\infty}\psi_{i,y+1}(y+1)y^2\mu + \sum_{y=0}^{\infty}\psi_{i,y-1}y^2\rho_k(i) \\
&\quad - \textstyle\sum_{y=0}^{\infty}\psi_{i,y}(y^3\mu + y^2\rho_k(i) - y^2A_{i,i}) \\
&= \textstyle\sum_{j\in V} A_{j,i}\Lambda^k_{2,j} + \sum_{y=0}^{\infty}\psi_{i,y+1}(y+1)y^2\mu - \sum_{y=0}^{\infty}\psi_{i,y}y^3\mu \\
&\quad + \textstyle\sum_{y=0}^{\infty}\psi_{i,y-1}\rho_k(i)y^2 - \sum_{y=0}^{\infty}\psi_{i,y}y^2\rho_k(i) \\
&= \textstyle\sum_{j\in V} A_{j,i}\Lambda^k_{2,j} + \sum_{y=0}^{\infty}\psi_{i,y+1}(y+1)^3\mu - \sum_{y=0}^{\infty}\psi_{i,y+1}(2y+1)(y+1)\mu \\
&\quad - \textstyle\sum_{y=0}^{\infty}\psi_{i,y}y^3\mu + \sum_{y=0}^{\infty}\psi_{i,y-1}\rho_k(i)(y-1)^2 + \sum_{y=0}^{\infty}\psi_{i,y-1}(2y-1)\rho_k(i) \\
&\quad - \textstyle\sum_{y=0}^{\infty}\psi_{i,y}y^2\rho_k(i) \\
&= \textstyle\sum_{j\in V} A_{j,i}\Lambda^k_{2,j} + \Lambda^k_{3,i}\mu - \sum_{y=0}^{\infty}\psi_{i,y+1}(2y+1)(y+1)\mu \\
&\quad -\Lambda^k_{3,i}\mu + \Lambda^k_{2,i}\rho_k(i) + \textstyle\sum_{y=0}^{\infty}\psi_{i,y-1}(2y-1)\rho_k(i) - \Lambda^k_{2,i}\rho_k(i)
\end{aligned}$$

$$
\begin{aligned}
&= \textstyle\sum_{j\in V} A_{j,i}\Lambda^k_{2,j} - \sum_{y=0}^{\infty} \psi_{i,y+1}(2y+1)(y+1)\mu \\
&\quad + \textstyle\sum_{y=0}^{\infty} \psi_{i,y-1}(2y-1)\rho_k(i) \\
&= \textstyle\sum_{j\in V} A_{j,i}\Lambda^k_{2,j} - 2\sum_{y=0}^{\infty} \psi_{i,y+1}(y+1)^2\mu + \sum_{y=0}^{\infty} \psi_{i,y+1}(y+1)\mu \\
&\quad +2\textstyle\sum_{y=0}^{\infty} \psi_{i,y-1}(y-1)\rho_k(i) + \sum_{y=0}^{\infty} \psi_{i,y-1}\rho_k(i) \\
&= \textstyle\sum_{j\in V} A_{j,i}\Lambda^k_{2,j} - 2\mu\Lambda^k_{2,i} + \mu\Lambda^k_{1,i} + 2\rho_k(i)\Lambda^k_{1,i} + \rho_k(i)\xi_i.
\end{aligned}
$$

Notice that $\sum_{i\in V} A_{j,i} = 0$ and $\Lambda^k_2 = \sum_{i\in V} \Lambda^k_{2,i}$. Summing the above over $i \in V$ yields

$$0 = \sum_{j\in V}\sum_{i\in V} A_{j,i}\Lambda^k_{2,j} - 2\mu\sum_{i\in V}\Lambda^k_{2,i} + \sum_{i\in V}[\mu\Lambda^k_{1,i} + 2\rho_k(i)\Lambda^k_{1,i} + \rho_k(i)\xi_i].$$

Equivalently,

$$0 = -2\Lambda^k_2\mu + \sum_{i\in V}[\Lambda^k_{1,i}(\mu + 2\rho_k(i)) + \xi_i\rho_k(i)],$$

which leads to (3.8). ∎

Proof of Lemma 4:

It is easy to show by induction that

$$\binom{s-1}{m-1} = \sum_{l=1}^{s-m+1}\binom{s-l-1}{m-2},$$

for any positive integers s, m and $s \geq m$. Hence

$$\sum_{l=1}^{s-m+1} g(m, s-l, l) = \left\{ \begin{array}{ll} \frac{1}{\binom{s-1}{m-1}}[\sum_{l=1}^{s-m+1}\binom{s-l-1}{m-2}] & \text{if} m > 1 \\ & \\ 0 + \cdots + 0 + 1 & \text{if} m = 1 \end{array} \right\} \equiv 1.$$

$\blacksquare$

Proof of Proposition 8:

Since $g(1, s+l, s) = 0$ for $s > 0$,

$$\gamma(s+l,s) = \sum_{m=2}^{s+1} \frac{m\mu_F g(m, s+l, s)}{\pi_{s+l}} \xi_{m,s+l} = \sum_{m=1}^{s} \frac{(m+1)\mu_F g(m+1, s+l, s)}{\pi_{s+l}} \xi_{m+1,s+l}.$$

Therefore

$$\begin{aligned}
\textstyle\sum_{l=1}^{S-s} \gamma(s+l,s)\pi_{s+l} &= \textstyle\sum_{l=1}^{S-s} \sum_{m=1}^{s} \frac{(m+1)\mu_F g(m+1,s+l,s)}{\pi_{s+l}} \xi_{m+1,s+l}\pi_{s+l} \\
&= \textstyle\sum_{l=1}^{S-s} \sum_{m=1}^{s} (m+1)\mu_F g(m+1, s+l, s)\xi_{m+1,s+l} \\
&= \textstyle\sum_{m=1}^{s} \sum_{l=1}^{S-s} (m+1)\mu_F \frac{\binom{s-1}{m-1}}{\binom{s+l-1}{m}} \frac{(\frac{\lambda}{\mu_F})^{m+1}}{(m+1)!} \binom{s+l-1}{m} p^{m+1}(1-p)^{s+l-m-1}\xi_{0,0} \\
&= \textstyle\sum_{m=1}^{s} \sum_{l=1}^{S-s} \frac{\lambda(\frac{\lambda}{\mu_F})^m}{m!} p^{m+1}(1-p)^{s+l-m-1}\binom{s-1}{m-1}\xi_{0,0} \\
&= \textstyle\sum_{l=1}^{S-s} \lambda p(1-p)^{l-1} \sum_{m=1}^{s} \xi_{m,s} \\
&= [\lambda \textstyle\sum_{l=1}^{S-s} p(1-p)^{l-1}]\pi_s.
\end{aligned}$$

Also, by lemma 4,

$$\begin{aligned}
\Gamma(s) &= \textstyle\sum_{l=1}^{s} \gamma(s, s-l) \\
&= \textstyle\sum_{l=1}^{s} \sum_{m=1}^{s-l+1} \frac{m\mu_F g(m,s,s-l)}{\pi_s} \xi_{m,s} \\
&= \textstyle\sum_{m=1}^{s} [\sum_{l=1}^{s-m+1} g(m, s, s-l)] \frac{m\mu_F}{\pi_s} \xi_{m,s} \\
&= \textstyle\sum_{m=1}^{s} \frac{m\mu_F}{\pi_s} \xi_{m,s}.
\end{aligned}$$

Hence, for $s > 0$,

$$
\begin{aligned}
\Gamma(s)\pi_s &= \textstyle\sum_{m=1}^{s} \frac{m\mu_F}{\pi_s}\xi_{m,s}\pi_s \\
&= \textstyle\sum_{m=1}^{s} m\mu_F\xi_{m,s} \\
&= \textstyle\sum_{m=1}^{s} \lambda\frac{(\frac{\lambda}{\mu_F})^{m-1}}{(m-1)!}\binom{s-1}{m-1}p^m(1-p)^{s-m}\xi_{0,0} \\
&= \textstyle\sum_{m=1}^{s} \lambda\frac{(\frac{\lambda}{\mu_F})^{m-1}}{(m-1)!}\left(\sum_{l=1}^{s-m+1}\binom{s-l-1}{m-2}\right)p^m(1-p)^{s-m}\xi_{0,0} \\
&= \textstyle\sum_{m=1}^{s}\left(\lambda\sum_{l=1}^{s-m+1} p(1-p)^{l-1}\xi_{m-1,s-l}\right) \\
&= \lambda\textstyle\sum_{l=1}^{s}\sum_{m=1}^{s-l+1} p(1-p)^{l-1}\xi_{m-1,s-l} \\
&= \lambda\textstyle\sum_{l=1}^{s} p(1-p)^{l-1}\sum_{m=1}^{s-l+1}\xi_{m-1,s-l} \\
&= \lambda\textstyle\sum_{l=1}^{s} p(1-p)^{l-1}\pi_{s-l}.
\end{aligned}
$$

Since $\xi_{0,s} = 0$ when $s > 0$ and $\pi_0 = \xi_{0,0}$, π_s satisfies the balance equations (3.10). ∎

Proof of Lemma 5:

Recall from the proof of Proposition 8 that $\Gamma(s) = \sum_{m=1}^{s} \frac{m\mu_F}{\pi_s}\xi_{m,s}$. Therefore, for all $s \in \{0, 1, \ldots, S\}$, we have

$$
\begin{aligned}
\sum_{i:n_0(i)=S-s} \xi_i A_{i,i} &= -\pi_s\left[\frac{\sum_{i:n_0(i)=S-s} \lambda \sum_{l=1}^{S-s} p(1-p)^{l-1} 1_{\{s\neq S\}}\xi_i + \sum_{m=1}^{s} m\mu_F \sum_{m(i)=m} \xi_i 1_{\{s\neq 0\}}}{\pi_s}\right] \\
&= -\pi_s\left[\lambda \sum_{l=1}^{S-s} p(1-p)^{l-1} 1_{\{s\neq S\}} \frac{\sum_{i:n_0(i)=S-s} \xi_i}{\pi_s} + \sum_{m=1}^{s} \frac{m\mu_F}{\pi_s} \xi_{m,s} 1_{\{s\neq 0\}}\right] \\
&= -\pi_s\left[\lambda \sum_{l=1}^{S-s} p(1-p)^{l-1} 1_{\{s\neq S\}} + \Gamma(s)\right] \\
&= \pi_s Q(s,s).
\end{aligned}
$$

To establish (3.12), we consider two cases.

Case 1: $d < s$ (a new order arrives at the FSL). In this case,

$$
\begin{aligned}
\sum_{i:n_0(i)=S-d} \sum_{j:n_0(j)=S-s} \xi_i A_{i,j} &= \sum_{i:n_0(i)=S-d} \sum_{j:n_0(j)=S-s} \xi_i \lambda p(1-p)^{s-d-1} 1_{\{\|(i,j)\|=1\}} \\
&= \sum_{i:n_0(i)=S-d} \xi_i \lambda p(1-p)^{s-d-1} \sum_{j:n_0(j)=S-s} 1_{\{\|(i,j)\|=1\}} \\
&= \sum_{i:n_0(i)=S-d} \xi_i \lambda p(1-p)^{s-d-1} \\
&= \pi_d \lambda p(1-p)^{s-d-1} \\
&= \pi_d Q(d,s),
\end{aligned}
$$

Case 2: $d > s$, (a replenishment arrives at the FSL). In this case,

$$\begin{aligned}
\sum_{i:n_0(i)=S-d}\sum_{j:n_0(j)=S-s}\xi_i A_{ij} &= \sum_{m=0}^{s}\sum_{\substack{i:m(i)=m+1\\ n_0(i)=d}}\sum_{\substack{j:m(j)=m\\ n_0(j)=s}}\xi_i A_{ij} \\
&= \sum_{m=0}^{s}\sum_{\substack{j:m(j)=m\\ n_0(j)=s}}\sum_{\substack{i:m(i)=m+1\\ n_0(i)=d}}\xi_i A_{ij} \\
\text{duetothereversibilityof} A &= \sum_{m=0}^{s}\sum_{\substack{j:m(j)=m\\ n_0(j)=s}}\sum_{\substack{i:m(i)=m+1\\ n_0(i)=d}}\xi_j A_{ji} \\
&= \sum_{m=1}^{s}\sum_{\substack{j:m(j)=m\\ n_0(j)=s}}\xi_j\sum_{\substack{i:m(i)=m+1\\ n_0(i)=d}}A_{ji} \\
&= \sum_{m=0}^{s}\sum_{\substack{j:m(j)=m\\ n_0(j)=s}}\xi_j\lambda p(1-p)^{d-s-1}\sum_{\substack{i:m(i)=m+1\\ n_0(i)=d}}1_{\{\|(i,j)\|=1\}} \\
&= \sum_{m=0}^{s}\xi_{m,s}\lambda p(1-p)^{d-s-1} \\
&= \sum_{m=0}^{s}\xi_{m+1,d}(m+1)\mu_F\frac{\binom{s-1}{m-1}}{\binom{d-1}{m}} \\
&= \pi_d\sum_{m=0}^{s}\xi_{m+1,d}\frac{(m+1)\mu_F g(m+1,d,s)}{\pi_d} \\
&= \pi_d\sum_{m=1}^{s+1}\xi_{m,d}\frac{m\mu_F g(m,d,s)}{\pi_d} \\
&= \begin{cases}\pi_d\sum_{m=2}^{s+1}\xi_{m,d}\frac{m\mu_F g(m,d,s)}{\pi_d} & \text{if} s>0\\[2ex] \pi_d\xi_{1,d}\frac{\mu_F}{\pi_1}\pi_1 & \text{if} s=0\end{cases} \\
&= \pi_d\gamma(d,s) \\
&= \pi_d Q(d,s).
\end{aligned}$$

The reversibility property is discussed in Chen *et al.* (2010).

∎

Proof of Proposition 9:

For fixed $s = 0, 1, \ldots, S$, sum (3.5) over $i \in \{V : n_0(i) = S - s\}$. This yields

$$0 = \sum_{i:n_0(i)=S-s} (\sum_{j\in V} A_{ji}\Lambda^k_{1,j}) - \mu\tilde{\Lambda}^k_{1,s} + \pi_s\tilde{\rho}_k(s).$$

By assumption $\psi^k_{y|j}$ is a constant over all j if $n_0(j) \equiv S - d$ for some d fixed,

$$\psi^k_{y|j}\pi_d = \sum_{i:n_0(i)=S-d} \psi^k_{y|i}\xi_i = \sum_{i:n_0(i)=S-d} \psi^k_{i,y},$$

for any $j \in \{i \in V, n_0(i) = S - d\}$. Let $\psi^k_{y|j} = \psi^k_{y|n_0(j)=S-d}$ be the probability of y emergency orders of size k conditioned on d units outstanding at the FSL. Then

$$\begin{aligned}
& \textstyle\sum_{i:n_0(i)=S-s}(\sum_{j\in V} A_{ji}\Lambda^k_{1,j}) \\
= \quad & \textstyle\sum_{i:n_0(i)=S-s}(\sum_{j\in V} \sum_{y=0}^{\infty} A_{ji}y\psi_{j,y}) \\
= \quad & \textstyle\sum_{i:n_0(i)=S-s}(\sum_{j\in V} \xi_j \sum_{y=0}^{\infty} \psi^k_{y|j}A_{ji}y) \\
= \quad & \textstyle\sum_{j\in V} \sum_{y=0}^{\infty} \sum_{i:n_0(i)=S-s} \xi_j\psi^k_{y|j}A_{ji}y \\
= \quad & \textstyle\sum_{d=0}^{S} \sum_{j:n_0(j)=S-d} \sum_{y=0}^{\infty} \sum_{i:n_0(i)=S-s} \psi^k_{y|j}y\xi_jA_{ji} \\
= \quad & \textstyle\sum_{d=0}^{S} \sum_{j:n_0(j)=S-d} \sum_{y=0}^{\infty} y\psi^k_{y|n_0(j)=S-d} \sum_{i:n_0(i)=S-s} \xi_jA_{ji}
\end{aligned}$$

$$= \quad \sum_{d=0}^{S} \sum_{y=0}^{\infty} y\psi^{k}_{y|n_0(j)=S-d} \sum_{j:n_0(j)=S-d} \sum_{i:n_0(i)=S-s} \xi_j A_{ji}$$

$$= \quad \sum_{d=0}^{S} \sum_{y=0}^{\infty} y\psi^{k}_{y|n_0(j)=S-d} \pi_d Q(d,s) (\text{Lemma.5})$$

$$= \quad \sum_{d=0}^{S} \tilde{\Lambda}^{k}_{1,d} Q(d,s).$$

Thus

$$0 = \sum_{d=0}^{S} \tilde{\Lambda}^{k}_{1,d} Q(d,s) - \mu\tilde{\Lambda}^{k}_{1,s} + \pi_s \tilde{\rho}_k(s).$$

Note that $Q - \mu I$ is non-singular since Q is also a generator of a continuous time Markov chain by construction. Therefore

$$(\tilde{\Lambda}^{k}_{1,s})_{1\times S+1} = (\pi_s \tilde{\rho}_k(s))_{1\times S+1} (Q - \mu I_{S+1\times S+1})^{-1}.$$

When calculating the second moment, observe that

$$\sum_{j\in V} [\Lambda^{k}_{1,j}(\mu + 2\rho_k(j)) + \xi_j \rho_k(j)]$$

$$= \quad \sum_{d=0}^{S} \sum_{j:n_0(j)=S-d} [\Lambda^{k}_{1,j}(\mu + 2\tilde{\rho}_k(d)) + \xi_j \tilde{\rho}_k(d)]$$

$$= \quad \sum_{d=0}^{S} [(\mu + 2\tilde{\rho}_k(d)) \sum_{j:n_0(j)=S-d} \Lambda^{k}_{1,j} + \pi_d \tilde{\rho}_k(d)]$$

$$= \quad \sum_{d=0}^{S} [\tilde{\Lambda}^{k}_{1,d}(\mu + 2\tilde{\rho}_k(d)) + \pi_d \tilde{\rho}_k(d)].$$

Therefore

$$\Lambda^{k}_{2} = \frac{\sum_{d=0}^{S} [\tilde{\Lambda}^{k}_{1,d}(\mu + 2\tilde{\rho}_k(d)) + \pi_d \tilde{\rho}_k(d)]}{2\mu}.$$

■

Proof of Corollary 3:

Since $Y_{k_1k_2,t} = Y_{k_1t} + Y_{k_2t}$,

$$\tilde{\Lambda}^{k_1k_2}_{1,d} = \tilde{\Lambda}^{k_1}_{1,d} + \tilde{\Lambda}^{k_2}_{1,d}$$

and

$$\Lambda^{k_1k_2}_1 = \Lambda^{k_1}_1 + \Lambda^{k_2}_1 = \sum_{d=0}^{S}(\tilde{\Lambda}^{k_1}_{1,d} + \tilde{\Lambda}^{k_2}_{1,d}).$$

The (k_1k_2)-system $N^{k_1k_2}_t = (N_t, Y_{k_1k_2,t})$ is a continuous time Markov chain with corresponding emergency order arrival rate $\rho_{k_1k_2}$. Applying the results from Proposition 7 for $N^{k_1k_2}_t$ yields

$$\Lambda^{k_1k_2}_2 = \sum_{i\in V}[\frac{\Lambda^{k_1k_2}_{1,i}(\mu + 2\rho_{k_1k_2}(i)) + \xi_i\rho_{k_1k_2}(i)}{2\mu}].$$

By repeating the process from the proof of Proposition 9, we have

$$\begin{aligned}
\Lambda^{k_1k_2}_2 &= \frac{\sum_{d=0}^{S}\tilde{\Lambda}^{k_1k_2}_{1,d}[\mu+2\tilde{\rho}_{k_1k_2}(d)]+\pi_d\tilde{\rho}_{k_1k_2}(d)}{2\mu}\\
&= \frac{\sum_{d=0}^{S}(\tilde{\Lambda}^{k_1}_{1,d}+\tilde{\Lambda}^{k_2}_{1,d})[\mu+2(\rho_{k_1}(d)+\rho_{k_2}(d))]+\pi_d(\rho_{k_1}(d)+\rho_{k_2}(d))}{2\mu}.
\end{aligned}$$

■

Proof of Proposition 10:

From the definition of the first moment of $\tilde{Z}$,

$$\mu = E(\tilde{Z}) = E(\tilde{Z}|\tilde{Z} = 0)p_0 + E(\tilde{Z}|\tilde{Z} > 0)(1 - p_0) = 0 + \frac{1}{1-\beta^{\alpha}}\alpha\frac{1-\beta}{\beta}(1 - p_0). \quad (3.26)$$

For the second moment of $\tilde{Z}$,

$$\begin{aligned}
\sigma^2 + \mu^2 = E(\tilde{Z}^2) &= E(\tilde{Z}^2|\tilde{Z} = 0)p_0 + E(\tilde{Z}^2|\tilde{Z} > 0)(1 - p_0) \\
&= 0 + \tfrac{1}{1-\beta^\alpha} E((1-\beta^\alpha)\tilde{Z}^2|\tilde{Z} > 0)(1 - p_0) \\
&= \tfrac{1}{1-\beta^\alpha}\left(\textstyle\sum_{\tilde{z}=0}^{\infty} \tilde{z}^2 \binom{\tilde{z}+\alpha-1}{\tilde{z}} \beta^\alpha (1-\beta)^{\tilde{z}}\right)(1 - p_0) \\
&= \tfrac{1}{1-\beta^\alpha}\left[\alpha\tfrac{1-\beta}{\beta^2} + \left(\alpha\tfrac{1-\beta}{\beta}\right)^2\right](1 - p_0) \\
&= \tfrac{1}{1-\beta^\alpha}\left(\alpha\tfrac{1-\beta}{\beta}\right)\left(\tfrac{1}{\beta} + \alpha\tfrac{1-\beta}{\beta}\right)(1 - p_0).
\end{aligned}$$

Therefore

$$\frac{\sigma^2 + \mu^2}{\mu} = \frac{1}{\beta} + \left(\alpha\frac{1-\beta}{\beta}\right) = \frac{1}{\beta}(1 + \alpha) - \alpha. \tag{3.27}$$

Consequently, β can be expressed as a function of α as given in (3.23). Substituting this expression for β in (3.26), we get

$$\frac{\mu}{(1 - p_0)} = \frac{1}{1 - \left(\frac{1+\alpha}{\frac{\sigma^2+\mu^2}{\mu}+\alpha}\right)^\alpha}\alpha\left(\frac{\frac{\sigma^2+\mu^2}{\mu} - 1}{1 + \alpha}\right). \tag{3.28}$$

Hence, α is a zero point of the function $f(r)$ defined as in (3.22). ■

3.E Partial Fill with Emergency Orders

We could use method similar to the approach for the complete fill case to get the mean and variance of emergency orders at the ESL for the partial fill case at the FSL. We restate only the notation which are changed in the proof.

It is easily seen that the infinitesimal generator for the Markov process N at the FSL is given by

$$A_{ij} \equiv \begin{cases} n_{k_{ij}}(i)\mu_F & \text{if}(i,j) \in V_R^2, \\ \lambda p_{k_{ij}} & \text{if}(i,j) \in V_C^2, n_0(j) > 0, \\ \lambda \overline{P}_{k_{ij}-1} & \text{if}(i,j) \in V_C^2, n_0(j) = 0, \\ -(m(i)\mu_F + \lambda 1_{\{n_0(i)\neq 0\}} & \text{if}\, j = i, \\ 0 & \text{otherwise,} \end{cases}$$

As shown in Chen *et al.*(2009), the stationary distribution of this process, N, is given by ξ_i for any state $i \in V$, where

$$\xi_i \equiv \frac{\left(\frac{\lambda p}{\mu_F(1-p)}\right)^{m(i)}}{\prod_{k=1}^{S}(n_k(i)!)G(S)} \frac{(1-p)^{S-n_0(i)}}{p^{1\{n_0(i)=0\}}}, \tag{3.29}$$

and $G(S) = \sum_{s=0}^{S}\sum_{m=0}^{s}\frac{(\frac{\lambda}{\mu})^m}{m!}p^{-1_{\{s=S\}}}f_{\rm NB}(s-m;m,p)$, $f_{\rm NB}(s-m;0,p) = 1\{s=0\}$ when $m = 0$.

Furthermore, the stationary distribution of the number of units on order, π_s ($s = 0, 1, \ldots, S$) is given by:

$$\pi_s = \frac{\sum_{m=0}^{s}\frac{\left(\frac{\lambda}{\mu_F}\right)^m}{p^{1\{s=S\}}}\frac{f_{NB}(s-m;m,p)}{m!}}{G(S)}. \tag{3.30}$$

3.E.1 The Mean and Second Moments of the Number of Emergency Orders

Similarly, let $(\psi^k_{i,y})$ denote the steady state distribution for the kth system in the partial fill case. Since the RSL-FSL system behaves as a backorder system, the number of orders in emergency resupply does not affect the evolution of the replenishment orders at the FSL. Therefore,

$$\xi_i = \sum_{y=0}^{\infty} \psi^k_{i,y},$$

where the vector ξ is given by (3.29). In addition, let $\psi^k_{y|i}$ denote the steady state distribution of y orders of size k at the ESL conditioned on state i at the FSL.

Recall that $\rho^j_k(i) \equiv 1_{\{n_0(i)=0, j\to i\}} \lambda p(1-p)^{n_0(j)+k-1}$ is the arrival rate for $Y_k(t)$ conditioned on the state of the FSL replenishment process.

Proposition 11 *For any k fixed, the $\Lambda^k_{1,i}$ satisfy the following equations:*

$$0 = \sum_{j\in V} A_{j,i}\Lambda^k_{1,j} - \mu\Lambda^k_{1,i} + \sum_{j\in V} \xi_j \rho^j_k(i), \text{for} i \in V. \tag{3.31}$$

That is, the row vector $\overrightarrow{\Lambda^k_{1\cdot}} = (\Lambda^k_{1,i})_{1\times|V|}$ has elements given by

$$\overrightarrow{\Lambda^k_{1\cdot}} = -(\sum_{j\in V} \xi_j \rho^j_k(i))_{1\times|V|} \times (A - \mu I_{|V|\times|V|})^{-1}, \tag{3.32}$$

where $|V|$ is the number of states of V and $I_{|V|\times|V|}$ is the $|V|$-dimensional identity matrix. In addition,

$$\Lambda^k_1 = \frac{\sum_{d=0}^{S} \pi_d \lambda p(1-p)^{S-d+k-1}}{\mu}. \tag{3.33}$$

$$\Lambda^k_2 = \frac{\sum_{i\in V}[\mu\Lambda^k_{1,i} + 2\sum_{j\in V}\Lambda^k_{1,j}\rho^j_k(i)] + \sum_{d=0}^{S} \pi_d \lambda p(1-p)^{S-d+k-1}}{2\mu}. \tag{3.34}$$

Proof: Let $\tilde{A}$ denote the infinitesimal generator of the continuous Markov chain for the k^{th} system. Then, for $(i, y) \in V \times N^+$,

$$
\begin{aligned}
\tilde{A}_{(i,y),(j,y)} &= A_{i,j} - \rho_k^i(j), \text{for } j \neq i; \\
\tilde{A}_{(i,y),(j,y+1)} &= \rho_k^i(j); \\
\tilde{A}_{(i,y),(i,y-1)} &= y\mu; \\
\tilde{A}_{(i,y),(i,y)} &= -(y\mu + \rho_k^i(i) - A_{i,i});
\end{aligned}
\tag{3.35}
$$

and the other elements of $\tilde{A}$ are 0. Define $\psi_{i,-1} = 0$. For any $(i, y) \in V \times N^+$, by the balance equation and substituting for (3.35), we get

$$
\begin{aligned}
0 &= \textstyle\sum_{j\neq i}(\psi_{j,y}\tilde{A}_{(j,y),(i,y)} + \psi_{j,y-1}\tilde{A}_{(j,y-1),(i,y)}) \\
&\quad +\psi_{i,y+1}\tilde{A}_{(i,y+1),(i,y)} + \psi_{i,y-1}\tilde{A}_{(i,y-1),(i,y)} + \psi_{i,y}\tilde{A}_{(i,y),(i,y)} \qquad (3.36) \\
&= \textstyle\sum_{j\neq i}(\psi_{j,y}(A_{j,i} - \rho_k^j(i)) + \psi_{j,y-1}\rho_k^j(i)) + \psi_{i,y+1}(y+1)\mu + \psi_{i,y-1}\rho_k^i(i) \\
&\quad -\psi_{i,y}(y\mu + \rho_k^i(i) - A_{i,i}).
\end{aligned}
$$

Since Λ_1^k and Λ_2^k are finite, multiplying (3.36) by y and summing over y, we get

$$\begin{aligned}
0 &= \textstyle\sum_{y=0}^{\infty}\sum_{j\neq i}\psi_{j,y}y(A_{j,i}-\rho_k^j(i)) + \sum_{y=0}^{\infty}\sum_{j\neq i}\psi_{j,y-1}y\rho_k^j(i) \\
&\quad \textstyle + \sum_{y=0}^{\infty}\psi_{i,y+1}(y+1)y\mu + \sum_{y=0}^{\infty}\psi_{i,y-1}m\rho_k^i(i) - \sum_{y=0}^{\infty}\psi_{i,y}(y^2\mu + m\rho_k^i(i) - mA_{i,i}) \\
&= \textstyle\sum_{j\in i}(A_{j,i}-\rho_k^j(i))\sum_{y=0}^{\infty}\psi_{j,y}y + \sum_{j\in V}(\rho_k^j(i))\sum_{y=0}^{\infty}\psi_{j,y-1}(y-1) \\
&\quad \textstyle + \sum_{j\in V}(\rho_k^j(i))\sum_{y=0}^{\infty}\psi_{j,y-1} + \sum_{y=0}^{\infty}\psi_{i,y+1}(y+1)^2\mu - \sum_{y=0}^{\infty}\psi_{i,y+1}(y+1)\mu \\
&\quad \textstyle - \sum_{y=0}^{\infty}\psi_{i,y}y^2\mu \\
&= \textstyle\sum_{j\in V}A_{j,i}\Lambda_{1,j}^k - \Lambda_{1,i}^k\mu + \sum_{j\in V}\sum_{y=0}^{\infty}\psi_{j,y-1}\rho_k^j(i) \\
&= \textstyle\sum_{j\in V}A_{j,i}\Lambda_{1,j}^k - \Lambda_{1,i}^k\mu + \sum_{j\in V}\xi_j\rho_k^j(i),
\end{aligned}$$

as claimed in (3.31).

In matrix form,

$$(\Lambda_{1,i}^k)_{1\times|V|}\cdot(A-\mu I_{|V|\times|V|}) = -(\sum_{j\in V}\xi_j\rho_k^j(i))_{1\times|V|}.$$

We claim $A-\mu I_{|V|\times|V|}$ is always non-singular. The proof can be adapted from p.407 where $A+I$ is shown to be non-singular. This leads to (3.32).

In addition, since $\sum_{i\in V}A_{j,i}=0$ and V is a finite state space, summing (3.31) for $i \in V$ yields:

$$
\begin{aligned}
0 &= \sum_{i\in V}\sum_{j\in V} A_{j,i}\Lambda^k_{1,j} - \sum_{i\in V}\Lambda^k_{1,i}\mu + \sum_{i\in V}\sum_{j\in V}\xi_j\rho^j_k(i) \\
&= -\Lambda^k_1\mu + \sum_{i\in V}\sum_{j\in V}\xi_j 1_{\{n_0(i)=0,j\to i\}}\lambda p(1-p)^{n_0(j)+k-1} \\
&= -\Lambda^k_1\mu + \sum_{i:n_0(i)=0}\sum_{j\in V:j\to i}\xi_j\lambda p(1-p)^{n_0(j)+k-1} \\
&= -\Lambda^k_1\mu + \sum_{d=0}^{S}\Big(\sum_{i:n_0(i)=0}\sum_{n_0(j)=S-d,j\to i}\xi_j\Big)\lambda p(1-p)^{S-d+k-1} \\
&= -\Lambda^k_1\mu + \sum_{d=0}^{S}\pi_d\lambda p(1-p)^{S-d+k-1}.
\end{aligned}
$$

The last equality is because $\{j : n_0(j) = S-d, j \to i, n_0(i) = 0\} = \{j : n_0(j) = S-d\}$. This leads to (3.33).

For Λ^k_2, the idea is similar. Since Λ^k_2, Λ^k_3 are finite, by multiplying (3.36) by y^2 and summing over y, we get:

$$
\begin{aligned}
0 &= \sum_{y=0}^{\infty}\sum_{j\neq i}\psi_{j,y}y^2(A_{j,i}-\rho^j_k(i)) + \sum_{y=0}^{\infty}\sum_{j\neq i}\psi_{j,y-1}y^2\rho^j_k(i) + \sum_{y=0}^{\infty}\psi_{i,y+1}(y+1)y^2\mu \\
&\quad + \sum_{y=0}^{\infty}\psi_{i,y-1}y^2\rho^i_k(i) - \sum_{y=0}^{\infty}\psi_{i,y}(y^3\mu + y^2\rho^i_k(i) - y^2A_{i,i}) \\
&= \sum_{j\in V}A_{j,i}\Lambda^k_{2,j} + \sum_{y=0}^{\infty}\psi_{i,y+1}(y+1)y^2\mu - \sum_{y=0}^{\infty}\psi_{i,y}y^3\mu \\
&\quad + \sum_{j\in V}\sum_{y=0}^{\infty}\psi_{j,y-1}\rho^j_k(i)2(y-1) + 2\sum_{y=0}^{\infty}\psi_{j,y-1}\rho^j_k(i) \\
&= \sum_{j\in V}A_{j,i}\Lambda^k_{2,j} - 2\mu\Lambda^k_{2,i} + \mu\Lambda^k_{1,i} + 2\sum_{j\in V}\Lambda^k_{1,j}\rho^j_k(i) + 2\sum_{j\in V}\xi_j\rho^j_k(i).
\end{aligned}
$$

Notice that $\sum_{i\in V} A_{j,i} = 0$, $\Lambda_2^k = \sum_{i\in V} \Lambda_{2,i}^k$ and $\sum_{i\in V}\sum_{j\in V} \xi_j \rho_k^j(i) = \sum_{d=0}^{S} \pi_d \lambda p(1-p)^{S-d+k-1}$. Summing the above over $i \in V$ yields

$$0 = \sum_{j\in V}\sum_{i\in V} A_{j,i}\Lambda_{2,j}^k - 2\mu \sum_{i\in V} \Lambda_{2,i}^k + \sum_{i\in V}[\mu\Lambda_{1,i}^k + 2\sum_{j\in V}\Lambda_{1,j}^k \rho_k^j(i) + 2\sum_{j\in V}\xi_j\rho_k^j(i)].$$

That is,

$$0 = -2\Lambda_2^k\mu + \sum_{i\in V}[\mu\Lambda_{1,i}^k + 2\sum_{j\in V}\Lambda_{1,j}^k\rho_k^j(i)] + \sum_{d=0}^{S}\pi_d\lambda p(1-p)^{S-d+k-1},$$

which leads to (3.34). ∎

Similarly, when S is small enough then $(A-\mu I)^{-1}$ is readily computable and we can solve equations (3.31) for $\Lambda_{1,j}^k$ and Λ_1^k by (3.33). Further we could get Λ_2^k by (3.34) for any size k. However, $|V|$ grows rapidly as S increases so this exact approach is computationally intractable when the demand over a lead time is substantial and shortages are to be avoided. As in the complete fill case, for a more general approach, we would use an approximation to decrease the dimension of the state space from $|V|$ to $S+1$.

3.E.2 The Mean and Second Moment Approximation of the Number of the Emergency Orders

In this section, we again construct a Markov Chain to approximate the steady state distribution of the number of units on order at the FSL. Now it is for the partial fill case. Then, we show how to simplify the formulas in Proposition 11 using this approximating Markov chain.

3.E.3 Markov Chain Approximation for the Number of Units on Order State Transition at the FSL

We define the infinitesimal generator of the approximating Markov chain for the partial fill case by Q :

$$Q(s,s') \equiv \begin{cases} \lambda p^{1_{\{s'<S\}}}(1-p)^{s'-s-1} & 0 \le s < s' \le S, \\ \gamma(s,s') & 0 \le s' < s \le S, \\ -(\Gamma(s) + \lambda 1_{\{s \neq S\}}) & s = s' \in \{0,1,2,...,S\} \\ 0 & \text{otherwise.} \end{cases}$$

The definition for $\gamma(s,s')$ and $\Gamma(s)$ are the same as the complete fill case.

So this Markov chain's stationary distribution $\{\tilde{\pi}_s, s = 0,1,2,\ldots,S\}$ should satisfy the balance equations:

$$[\lambda + \Gamma(s)]\tilde{\pi}_s = \lambda \sum_{k=1}^{s} p^{1_{\{s<S\}}}(1-p)^{k-1}\tilde{\pi}_{s-k} + \sum_{k=1}^{S-s} \gamma(s+k,s)\tilde{\pi}_{s+k}, \text{for} s = 1,2,\ldots, \quad (3.37)$$

and $\sum_{s=0}^{S} \tilde{\pi}_s = 1$.

Proposition 12 *In the partial fill case, the stationary distribution stated π_s, given by (3.30), also satisfies this balance equation. That is, $\tilde{\pi}_s = \pi_s$ for $s = 0,1,2,\ldots,S$.*

Proof: Since $g(1,s,k) = 0$ for $s > 0$,

$$\gamma(s+k,s) = \sum_{m=1}^{s} \frac{(m+1)\mu g(m+1,s+k,s)}{\pi_{s+k}} \xi_{m+1,s+k}.$$

Therefore

$$
\begin{aligned}
& \sum_{k=1}^{S-s} \gamma(s+k,s)\pi_{s+k} \\
= \quad & \sum_{k=1}^{S-s} \sum_{m=1}^{s} \frac{(m+1)\mu g(m+1,s+k,s)}{\pi_{s+k}} \xi_{m+1,s+k}\pi_{s+k} \\
= \quad & \sum_{k=1}^{S-s} \sum_{m=1}^{s} (m+1)\mu g(m+1,s+k,s)\xi_{m+1,s+k} \\
= \quad & \sum_{m=1}^{s} \sum_{k=1}^{S-s} (m+1)\mu \frac{\binom{s-1}{m-1}}{\binom{s+k-1}{m}} \frac{(\frac{\lambda}{\mu})^{m+1}}{(m+1)!} \binom{s+k-1}{m} p^{m+1}(1-p)^{s+k-m-1} \frac{1}{p^{1_{\{s+k=S\}}}} \xi_{0,0} \\
= \quad & \sum_{m=1}^{s} \sum_{k=1}^{S-s} \frac{\lambda(\frac{\lambda}{\mu})^{m}}{m!} p^{m+1}(1-p)^{s+k-m-1} \binom{s-1}{m-1} \frac{1}{p^{1_{\{s+k=S\}}}} \xi_{0,0} \\
= \quad & [\sum_{k=1}^{S-s-1} \lambda p(1-p)^{k-1} + \lambda \frac{p(1-p)^{S-s-1}}{p}] \sum_{m=1}^{s} \xi_{m,s} \\
= \quad & \lambda\pi_s.
\end{aligned}
$$

Also,

$$\sum_{k=1}^{s-m+1} g(m,s,s-k) = 1$$

$$
\begin{aligned}
\Gamma(s) \quad = \quad & \sum_{k=1}^{s} \gamma(s,s-k) \\
= \quad & \sum_{k=1}^{s} \sum_{m=1}^{s-k+1} \frac{m\mu g(m,s,s-k)}{\pi_s} \xi_{m,s} \\
= \quad & \sum_{m=1}^{s} [\sum_{k=1}^{s-m+1} g(m,s,s-k)] \frac{m\mu}{\pi_s} \xi_{m,s} \\
= \quad & \sum_{m=1}^{s} \frac{m\mu}{\pi_s} \xi_{m,s}.
\end{aligned}
$$

Hence, for $s > 0$,

$$\begin{aligned}
\Gamma(s)\pi_s &= \textstyle\sum_{m=1}^{s} \frac{m\mu}{\pi_s}\xi_{m,s}\pi_s \\
&= \textstyle\sum_{m=1}^{s} m\mu\xi_{m,s} \\
&= \textstyle\sum_{m=1}^{s} \lambda\frac{(\frac{\lambda}{\mu})^{m-1}}{(m-1)!}\binom{s-1}{m-1}p^m(1-p)^{s-m}\frac{1}{p^{1_{\{s=S\}}}}\xi_{0,0} \\
&= \textstyle\sum_{m=1}^{s} \lambda\frac{(\frac{\lambda}{\mu})^{m-1}}{(m-1)!}(\sum_{k=1}^{s-m+1}\binom{s-k-1}{m-2})p^m(1-p)^{s-m}\frac{1}{p^{1_{\{s=S\}}}}\xi_{0,0} \\
&= \textstyle\sum_{m=1}^{s}(\lambda\sum_{k=1}^{s-m+1}\frac{p(1-p)^{k-1}}{p^{1_{\{s=S\}}}}\xi_{m-1,s-k}) \\
&= \textstyle\lambda\sum_{k=1}^{s}\sum_{m=1}^{s-k+1}\frac{p(1-p)^{k-1}}{p^{1_{\{s=S\}}}}\xi_{m-1,s-k} \\
&= \textstyle\lambda\sum_{k=1}^{s}\frac{p(1-p)^{k-1}}{p^{1_{\{s=S\}}}}\sum_{m=1}^{s-k+1}\xi_{m-1,s-k} \\
&= \textstyle\lambda\sum_{k=1}^{s} p^{1_{\{s<S\}}}(1-p)^{k-1}\pi_{s-k}.
\end{aligned}$$

The last equality comes from $\xi_{0,s} = 0$ if $s > 0$, and $\pi_0 = \xi_{0,0}$.

So π_s satisfies the balance equation (3.37). ∎

The generators A and Q are defined on different state spaces. We relate these two generators in the following lemma.

Lemma 6 *For each $i \in V$ and $n(i) = (n_1(i), n_2(i), \ldots, n_s(i))$ at the FSL when demands occur according to a stuttering Poisson process (partial fills), for fixed $s = 0, 1, \ldots, S$,*

$$\sum_{i:n_0(i)=S-s} \xi_i A_{i,i} = \pi_s Q(s, s),$$

and, for fixed $d \in \{0, 1, \ldots, S\}$ *and* $d \neq s$,

$$\sum_{i:n_0(i)=S-d} [\sum_{j:n_0(j)=S-s} \xi_i A_{i,j}] = \pi_d Q(d, s). \tag{3.38}$$

Proof: For $\forall s$ fixed,

$$\begin{aligned}
\textstyle\sum_{i:n_0(i)=S-s} \xi_i A_{i,i} &= -\pi_s[\frac{\sum_{i:n_0(i)=S-s} \lambda[\sum_{k=1}^{S-s-1} p(1-p)^{k-1}+(1-p)^{S-s-1}]1_{\{s\neq S\}}\xi_i+\sum_{m=1}^{s} m\mu \sum_{m(i)=m} \xi_i 1_{\{s\neq 0\}}}{\pi_s}] \\
&= -\pi_s[\lambda \frac{\sum_{i:n_0(i)=S-s} \xi_i}{\pi_s} 1_{\{s\neq S\}} + \textstyle\sum_{m=1}^{s} \frac{m\mu}{\pi_s} \xi_{m,s} 1_{\{s\neq 0\}}] \\
&= -\pi_s[\lambda 1_{\{s\neq S\}} + \Gamma(s)] \\
&= \pi_s Q(s, s).
\end{aligned}$$

To establish (3.38), we consider two cases.

Case 1: $d < s$, new order arrives at FSL:

$$\begin{aligned}
\textstyle\sum_{i:n_0(i)=S-d} \sum_{j:n_0(j)=S-s} \xi_i A_{i,j} &= \textstyle\sum_{i:n_0(i)=S-d} \sum_{j:n_0(j)=S-s} \xi_i \lambda p^{1_{\{s<S\}}} (1-p)^{s-d-1} 1_{\{\|(i,j)\|=1\}} \\
&= \textstyle\sum_{i:n_0(i)=S-d} \xi_i \lambda p^{1_{\{s<S\}}} (1-p)^{s-d-1} \sum_{j:n_0(j)=S-s} 1_{\{\|(i,j)\|=1\}} \\
&= \textstyle\sum_{i:n_0(i)=S-d} \xi_i \lambda p^{1_{\{s<S\}}} (1-p)^{s-d-1} \\
&= \pi_d \lambda p^{1_{\{s<S\}}} (1-p)^{s-d-1} \\
&= \pi_d Q(d, s),
\end{aligned}$$

Case 2: $d > s$, replenishment arrives at FSL:

$$\begin{aligned}
\sum_{i:n_0(i)=S-d}\sum_{j:n_0(j)=S-s}\xi_i A_{ij} &= \sum_{m=1}^{s}\sum_{\substack{i:m(i)=m+1\\ n_0(i)=d}}\sum_{\substack{j:m(j)=m\\ n_0(j)=s}}\xi_i A_{ij}\\
&= \sum_{m=1}^{s}\sum_{\substack{j:m(j)=m\\ n_0(j)=s}}\sum_{\substack{i:m(i)=m+1\\ n_0(i)=d}}\xi_i A_{ij}\\
&\quad \text{duetothereversibility}\\
&= \sum_{m=1}^{s}\sum_{\substack{j:m(j)=m\\ n_0(j)=s}}\sum_{\substack{i:m(i)=m+1\\ n_0(i)=d}}\xi_j A_{ji}\\
&= \sum_{m=1}^{s}\sum_{\substack{j:m(j)=m\\ n_0(j)=s}}\xi_j\sum_{\substack{i:m(i)=m+1\\ n_0(i)=d}}A_{ji}\\
&= \sum_{m=1}^{s}\sum_{\substack{j:m(j)=m\\ n_0(j)=s}}\xi_j\lambda p^{1_{\{d<S\}}}(1-p)^{d-s-1}\sum_{\substack{i:m(i)=m+1\\ n_0(i)=d}}1_{\{\|(i,j)\|=1\}}\\
&= \sum_{m=1}^{s}\xi_{m,s}\lambda p^{1_{\{d<S\}}}(1-p)^{d-s-1}\\
&= \sum_{m=1}^{s}\xi_{m+1,d}(m+1)\mu\frac{\binom{s-1}{m-1}}{\binom{d-1}{m}}\\
&= \pi_d\sum_{m=1}^{s}\xi_{m+1,d}\frac{(m+1)\mu g(m+1,s,d-s)}{\pi_d}\\
&= \pi_d\gamma(d,s)\\
&= \pi_d Q(d,s).
\end{aligned}$$

∎

In summary, we have defined the generator Q of a Markov chain on the number of units on order at the FSL for the partial fill case that yields steady state distribution as the true lost sales model. We also derived an equivalence (Lemma 6) that will be used in the next subsection to collapse the state space.

3.E.4 The Mean and Second Moment Approximation

Proposition 13 *Suppose $\psi^k_{y|i} = \psi^k_{y|i'}$, $\forall i, i' \in \{j \in V : n_0(j) \equiv n_0 j'\}$. Then for $d = 0, 1, \ldots, S$, $\tilde{\Lambda}^k_{1,d}$ will satisfy*

$$0 \equiv \sum_{d=0}^{S} \tilde{\Lambda}^k_{1,d} Q(d,s) - \mu \tilde{\Lambda}^k_{1,s} + \sum_{d=0}^{S} \pi_d 1_{\{s=S\}} \lambda p (1-p)^{S-d+k-1},$$

that is,

$$(\tilde{\Lambda}^k_{1,s})_{1\times S+1} = -(1_{\{s=S\}} \sum_{d=0}^{S} \pi_d \lambda p (1-p)^{S-d+k-1})_{1\times S+1} [(Q - \mu I_{S+1\times S+1})^{-1}], \tag{3.39}$$

where $I_{S+1\times S+1}$ is the $S+1$-dimensional identity matrix. Furthermore,

$$\Lambda^k_2 = \frac{\sum_{d=0}^{S} \tilde{\Lambda}^k_{1,d}\mu + \sum_{d=0}^{S}(2\tilde{\Lambda}^k_{1,d} + \pi_d)\lambda p (1-p)^{S-d+k-1}}{2\mu}. \tag{3.40}$$

Proof: For fixed $s = 0, 1, \ldots, S$, sum over $i \in \{V : n_0(i) = S - s\}$ in equation (3.31). We could get

$$0 = \sum_{i:n_0(i)=S-s} (\sum_{j\in V} A_{j,i} \Lambda^k_{1,i}) - \mu \tilde{\Lambda}^k_{1,s} + \sum_{i:n_0(i)=S-s} \sum_{j\in V} \xi_j \rho^j_k(i).$$

Since $\psi^k_{y|j}$ is a constant if $n_0(j) \equiv S - d$ for some d fixed,

$$\psi^k_{y|j}\pi_d = \sum_{i:n_0(i)=S-d} \psi^k_{y|i}\xi_i = \sum_{i:n_0(i)=S-d} \psi^k_{i,y},$$

for any $j \in \{i \in V, n_0(i) = S - d\}$. We could write $\psi^k_{y|j} = \psi^k_{y|n_0(j)=S-d}$ as the probability conditioned on d units outstanding at the FSL,

$$\begin{aligned}
\textstyle\sum_{i:n_0(i)=S-s}(\sum_{j\in V} A_{j,i}\Lambda^k_{1,i}) &= \textstyle\sum_{i:n_0(i)=S-s}[\sum_{j\in V}\xi_j \sum_{y=0}^{\infty}\psi^k_{y|j}A_{ji}y] \\
&= \textstyle\sum_{d=0}^{S}\sum_{j:n_0(j)=S-d}\sum_{y=0}^{\infty}\sum_{i:n_0(i)=S-s}\psi^k_{y|j}m\xi_j A_{ji} \\
&= \textstyle\sum_{d=0}^{S}\sum_{j:n_0(j)=S-d}\sum_{y=0}^{\infty} y\psi^k_{y|n_0(j)=S-d}\sum_{i:n_0(i)=S-s}\xi_j A_{ji} \\
&= \textstyle\sum_{d=0}^{S}\sum_{y=0}^{\infty} y\psi^k_{y|n_0(j)=S-d}\sum_{j:n_0(j)=S-d}\sum_{i:n_0(i)=S-s}\xi_j A_{ji} \\
&= \textstyle\sum_{d=0}^{S}\sum_{y=0}^{\infty} y\psi^k_{y|n_0(j)=S-d}\pi_d Q(d,s)(\text{Lemma.6}) \\
&= \textstyle\sum_{d=0}^{S}\tilde{\Lambda}^k_{1,d}Q(d,s),
\end{aligned}$$

and

$$\begin{aligned}
\textstyle\sum_{i:n_0(i)=S-s}\sum_{j\in V}\xi_j\rho^j_k(i) &= \textstyle\sum_{i:n_0(i)=S-s}\sum_{j\in V}\xi_j 1_{\{n_0(i)=0,j\to i\}}\lambda p(1-p)^{n_0(j)+k-1} \\
&= \textstyle 1_{\{s=S\}}\sum_{j\in V:j\to i,n_0(i)=0}\xi_j\lambda p(1-p)^{n_0(j)+k-1} \\
&= \textstyle 1_{\{s=S\}}\sum_{d=0}^{S}\pi_d\lambda p(1-p)^{S-d+k-1}.
\end{aligned}$$

Thus,

$$0 = \sum_{d=0}^{S}\tilde{\Lambda}^k_{1,d}Q(d,s) - \mu\tilde{\Lambda}^k_{1,s} + 1_{\{s=S\}}\sum_{d=0}^{S}\pi_d\lambda p(1-p)^{S-d+k-1}.$$

Here $Q - \mu I$ is still non-singular since Q is also a generator of a continuous Markov chain by construction. This leads to (3.39).

For the second moment calculation, observe that

$$\begin{aligned}
\sum_{i\in V}[\mu\Lambda_{1,i}^{k} + 2\sum_{j\in V}\Lambda_{1,j}^{k}\rho_k^j(i)] &= \sum_{d=0}^{S}\sum_{i:n_0(i)=S-d}[\mu\Lambda_{1,i}^{k} + 2\sum_{j\in V}\Lambda_{1,j}^{k}\rho_k^j(i)] \\
&= \sum_{d=0}^{S}\mu\tilde{\Lambda}_{1,d}^{k} + 2\sum_{d=0}^{S}\sum_{i:n_0(i)=S-d}\sum_{j\in V}\Lambda_{1,j}^{k}\rho_k^j(i) \\
&= \sum_{d=0}^{S}\tilde{\Lambda}_{1,d}^{k}\mu \\
&\quad +2\sum_{i:n_0(i)=0}\sum_{\{j\in V:j\to i,n_0(i)=0\}}\Lambda_{1,j}^{k}\lambda p(1-p)^{n_0(j)+k-1} \\
&= \sum_{d=0}^{S}\tilde{\Lambda}_{1,d}^{k}\mu + 2\sum_{d=0}^{S}\tilde{\Lambda}_{1,d}^{k}\lambda p(1-p)^{S-d+k-1}
\end{aligned}$$

This leads to (3.40) ∎

BIBLIOGRAPHY

[1] Alfredsson, P. and J. Verrijdt. 1999. Modeling Emergency Supply Flexibility in a Two-Echelon Inventory System. *Management Science*, 45, No. 10, 1416-1431.

[2] Allen, S.G. and D.A. D'Esopo. 1968. An Ordering Policy for Stock Items When Delivery can be Expedited. *Operations Research*, 16, 880-883.

[3] Axsäer, S.. 1990. Modeling emergency lateral transshipments in inventory systems. *Management Science*, 36, 1329-1338.

[4] Axsäter, S.. 2003. A New Decision Rule for Lateral Transshipments in Inventory Systems. *Management Science*, 49, No. 9, 1168-1179.

[5] Axsäter, S.. 2005. A heuristic for triggering emergency orders in an inventory system. *European Journal of Operational Research*, 176 (2007) 880C891.

[6] Axsäter, S.. 2006. *Inventory Control*, 2nd ed., Springer, New York.

[7] Archibald, T.W., S.A. Sassen, and L.C. Thomas. 1997. An optimal policy for a two depot inventory problem with stock transfer. *Management Science*, 43, 173C183.

[8] Baranikin, E.W., 1961. A Delivery-Lag Inventory Model with an Emergency Provision. *Naval Res. Logis. Quart*, 8 , 285-311.

[9] Blumenfeld, D.E., R.W. Hall, and W.C. Jordan. 1985. Trade-off between Freight Expediting and Safety Stock Inventory Costs. *J. Bus. Logist.*, 6, 79-100.

[10] Bulinskaya, E.. 1964 Some Results Concerning Optimal Inventory Policies. *Theory Prob. Appli.*, 9, 502-507.

[11] Chen, J., P. Jackson, and J. Muckstadt. Exact Analysis of a Lost Sales Model under Stuttering Poisson Demand. Technical Note, *Operations Research* (to appear 2010). Technical Report No. 1474, School of Operations Research and Information Engineering, Cornell University.

[12] Chiang, C.. 2002. Ordering Policies for Inventory Systems with Expediting. *The Journal of the Operational Research Society*, Vol 53, No.11, 1291-1295.

[13] Chiang, C. and G.J. Gutierrez. 1996. A Periodic Review Inventory System with Two Supply Modes. *Eur. J. Oper. Res.*, 94, 527-547.

[14] Chiang, C. and G.J. Gutierrez. 1998. Optimal Control Policies for a Periodic Review Inventory System with Emergency Orders. *Naval Research Logistics*, 45, 187-204.

[15] Dada, M.. 1992. A Two-Echelon Inventory System with Priority Shipments, *Management Science*. 38, 1140-1153.

[16] Daniel, K.H.. 1962. A Delivery-Lag Inventory Model with Emergency. H.E. Scarf, D.M. Gilford, M.W. Shelly, eds. *Multistage Inventory Models and Techniques*. Stanford University Press, Stanford, CA.

[17] Das, C.. 1975. Supply and Redistribution Rules for Two-Location Inventory Systems: One Period Analysis, *Management Science*, 21, 765-779.

[18] Downs, B., R. Metters, and J. Semple. 2001. Managing Inventory with Multiple Products, Lags in Delivery, Resource Constraints, and Lost Sales: a Mathematical Programming Approach. *Management Science*, 47(3), 464-479.

[19] Feeney, G.J. and C.C. Sherbrooke. 1966. The (S-1,S) Inventory Policy under Compound Poisson Demand. *Management Science*, 12, No. 5, Series A, Sciences., pp. 391-411.

[20] Feller,W.. 1971 *An Introduction to Probability Theory and Its Applications*, Vol.2, 2nd ed., Wiley, New York, pp. 19.

[21] Feng, Q., G. Gallego, S.P. Sethi, H. Yan, and H. Zhang. 2005. A Periodic Review Inventory Model with Three Consecutive Delivery Modes and Forecast Updates. *J. Optim. Theory Appl.*, 124, 137-155.

[22] Feng, Q., S.P. Sethi, H. Yan, and H. Zhang. 2006. Optimality and Non Optimality of the Base-stock Policy in Inventory Problems with Multiple Delivery Modes. *J. Indust. Management Optim.*, 2, 19-42.

[23] Feng, Q., S.P. Sethi, H. Yan, and H. Zhang. 2006. Are Base-Stock Policies Optimal in Inventory Problems with Multiple Delivery Modes? *Operations Research*, Technical Note., Vol. 54, No. 4, 801-807

[24] Fukuda, Y.. 1964. Optimal Policy for the Inventory Problem with Negotiable Leadtime. *Management Science*, 10, 609-708.

[25] Grahovac, J. and A. Chakravarty. 2001. Sharing and Lateral Transshipment of Inventory in a Supply Chain with Expensive, Low-Demand Items. *Management Science*, 47, 579-594.

[26] Gross, D. and A. Soriano. 1972. On the Economic Application of Airlift to Product Distribution and Its Impact on Inventory Levels. *Naval Res. Logis. Quart*, 19, 501-507.

[27] Hadley, G. and T.M. Whitin. 1963. *Analysis of Inventory Systems*. Pentice-Hall, Inc. Englewood Cliffs, New Jersey, pp. 204-212.

[28] Huggins, E. and T. Olsen. 2001. Inventory control with generalized expediting. Working paper, Fort Lewis College, Durango, CO.

[29] Huggins, E. and T. Olsen. 2003. Supply Chain Management with Guaranteed Delivery. *Management Science*, 49, No. 9, 1154-1167.

[30] Johansen, S.G. and A. Thorstenson. 1998. An inventory model with Poisson demands and emergency orders. *International Journal of Production Economics*, 56C57, 275C289.

[31] Kelly, F.P.. 1991. Loss Networks. *The Annals of Applied Probability*, Vol. 1, No. 3., pp. 319-378.

[32] Kranenburg, A. and G. van Houtum . 2009. A new partial pooling structure for spare parts networks. *International Journal of Production Economics*, 108, 221-227.

[33] Kukreja, A., D.M. Miller, and C.P. Schmidt. 2001. Stocking decisions for low-usage items in a multilocation inventory system. *Management Science*, 47, 1371C1383.

[34] Lawson, D.G. and E.L. Porteus. 2000. Multistage Inventory Management with Expediting. *Operations Research*, 48, 878-893.

[35] Lee, H.L., 1987. A multi-echelon inventory model for repairable items with emergency lateral transshipments. *Management Science*, 33, 1302C1316.

[36] Minner, S., 2003. Multiple-supplier inventory models in supply chain management: A review. *International Journal of Production Economics*, 81C82, 265C279.

[37] Minner, S., D.J. Robb, and E.A. Silver. 2003. An improved heuristic for deciding on emergency transshipments. European Journal of *Operational Research*, 148, 384C400.

[38] Moinzadeh, K. and PK. Aggarwal. 1997. An Information Based Multi-echolon Inventory System with Emergency Orders. *Operations Research*, 45, 694-701.

[39] Moinzadeh, K. and S. Nahmias. 1988. A Continuous Review Model for an Inventory System with Two Supply Modes. *Management Science*, 6, 761-773.

[40] Moinzadeh, K.. 1989. Operating Characteristics of the (S-1,S) Inventory System with Paritial Backorders and Constant Resupply Times. *Management Science*, 35, No. 4., pp. 472-477.

[41] Moinzadeh, K., and C.P. Schmidt. 1991. An (S-1,S) Inventory System with Emergency Orders. *Operations Research*, 39, No. 2. pp. 308-321.

[42] Muckstadt, J.. 2005. *Analysis and Algorithms for Service Parts Supply Chains*, Springer.

[43] Muckstadt, J., and L. J. Thomas. 1980. Are Multi- Echelon Inventory Methods Worth Implementing in Systems With Low-Demand-Rate Items? *Management Science*, 26, 483-494

[44] Muharremoğlu, A. and J. N. Tsitsiklis. 2003. Dynamic Leadtime Management in Supply Chains, working paper. Graduate School of Business, Columbia University, New York, NY 10027.

[45] Nahmias, S.. 1979. Simple Approximations for a Variety of Dynamic Leadtime Lost-Sales Inventory Models. *Operations Research*, 27(5), 904-924.

[46] Nahmias, S.. 2005. *Production and Operations Analysis*, McGraw-Hill, 5 edition.

[47] Neuts, F.. 1964. An Inventory Model with an Optional Time Lag. *SIAM J. Appl. Math.*, 12, 179-185.

[48] Olsson, F.. 2009. Optimal Policies for Inventory Systems with Lateral Transshipments, *Int. J.ProductionEconomics*, 118, 175-184.

[49] Paterson C., G. Kiesmüller, R. Teunter, and K. Glazebrook, 2009, Inventory models with lateral transshipments: A review, Beta Working Paper 287, Eindhoven University of Technology.

[50] Porteus, E.. 2002. *Foundations of Stochastic Inventory Theory*, Stanford University Press, first ed.

[51] Puterman, M.L.. 2005 *Markov Decision Processes Discrete Stochastic Dynamic Programming*, John Wiley&Sons, Inc.

[52] Reijnen, I.C., T. Tan, and G.J. van Houtum. 2009. Inventory Planning for spare parts networks with delivery time requirements. Working paper, Eindhoven University of Technology.

[53] Resnick, S.I., *Adventures in Stochastic Processes*, Birkhauser, Boston, (2005).

[54] Robinson, L. W.. 1990. Optimal and Approximate Policies in Multiperiod, Multilocation Inventory Models with Transshipments. *Operations Research*, 38, 278-295.

[55] Rosenshine, M. and D. Obee. 1976. Analysis of a Standing Order Inventory System with Emergency Orders. *Operations Research*, 24, 1143-1155.

[56] Rudi, N., S. Kapur, and D. F. Pyke. 2001. A Two-Location Inventory Model with Transshipment and Local Decision Making, *Management Science*, 47, 1668-1680.

[57] Sherbrooke, C.C. 1992. Multiechelon Inventory Systems with Lateral Supply. *Naval Research Logistics*, 39, 29-40.

[58] Smith, S.. 1977. Optimal Inventories for an (S-1,S) system with no backorders. *Management Science*, 23, 522-528.

[59] Tagaras, G. and M.A. Cohen, 1992. Pooling in two-location inventory systems with non-negligible replenishment lead times. *Management Science*, 38, 1067C1083.

[60] Tagaras, G. and D. Vlachos. 2001. A Periodic Review Inventory System with Emergency Replenishments. *Management Science*, 47, 415-429.

[61] Teunter, R. and D. Vlachos. 2001. An Inventory System with Periodic Regular Review and Flexible Emergency Review. *IIE Transactions*, 33, 625-635.

[62] Veinott, A. F. JR. 1966. The Status of Mathematical Inventory Theory. *Management Science*, 12, 745-777.

[63] Vidgren, S.. 2005. Evaluation of a Class of Lateral Transshipment Policies, Lund University.

[64] Vliegen, I.. 2009. *Integrated planning for service tools and spare parts for capital goods*, Ipskamp Drukkers.

[65] Wijk A. van, I. Adan, and G. van Houtum. 2009. Optimal lateral transshipment policy for a two location inventory problem. Working paper, Eindhoven University of Technology.

[66] Wong, H., G.J. Van Houtum, D. Cattrysse, and D. Van Oudheusden. 2006. Multi-item spare parts systems with lateral transshipment and waiting time constraints. *European Journal of Operational Research*, 171, 1071C1093.

[67] Wright, G.P. 1969. Optimal Ordering Policies for inventory systems with emergency orders. *Naval Res. Logist. Quart.* 15, 375-401.

[68] Wittmore, A.S. and S. Saunders. 1977. Optimal Inventory under Stochastic Demand with Two Supply Options. *SLAM J. Appl. Math.*, 32, 293-305.

[69] Zipkin, P.H.. 2000. *Foundations of Inventory Management*, McGraw-Hill.

[70] Zhang, V. 1996. Ordering policies for an inventory system with three supply modes. *Naval Research Logistics*, 43, 691-708.

CHAPTER 4

STOCK OPTIMIZATION IN EMERGENCY RESUPPLY NETWORKS UNDER STUTTERING POISSON DEMAND

Jie Chen, Peter L. Jackson, John A. Muckstadt

School of Operations Research and Information Engineering,

Cornell University, Ithaca, NY, USA

jc562@cornell.edu, pj16@cornell.edu, jam61@cornell.edu

Abstract: We consider a network in which field stocking locations (FSLs) manage multiple parts according to an (S-1,S) policy. Demand processes for the parts are assumed to be independent stuttering Poisson processes. Regular replenishments to an FSL occur from a regional stocking location (RSL) that has an unlimited supply of each part type. Demand in excess of supply at an FSL is routed to an emergency stocking location (ESL), which also employs an (S-1,S) policy to manage its inventory. Demand in excess of supply at the ESL is backordered. Lead time from the ESL to each FSL is assumed to be negligible compared to the RSL-ESL resupply time. In companion papers we have shown how to approximate the joint probability distributions of units on hand, units in regular resupply, and units in emergency resupply. In this paper, we focus on the problem of determining the stock levels at the FSLs and ESL across all part numbers that minimize backorder, and emergency resupply costs subject to an inventory investment budget constraint. The problem is shown to be a non-convex integer programming problem, and we explore a collection of heuristics for solving the optimization problem.

4.1 Introduction and Literature Review

4.1.1 Introduction

In this paper we study a system containing a regional stocking location (RSL), which serves two types of facilities: a set of N field service locations (FSL) and an emergency stocking location (ESL). Each location stocks multiple part types, which are used by technical service representatives who make visits to customer sites to repair equipment. In our model, we use a stuttering Poisson process to represent the demand process for each of I part types at each FSL. We have employed this model since the variance to mean ratio of the demand at the FSLs is greater than 1 in systems that we have examined. We assume the inventory control policy followed at each location is an (s-1,s) or order-up-to, policy. Again, this policy type is the one used in applications we have studied. The system we will examine works as follows. When a customer order occurs, if the on hand inventory at the FSL is sufficient to satisfy the entire customer demand, we fulfill this order from the FSL stock, and immediately place a regular replenishment order of the same order size on the RSL. Whenever a customer's demand exceeds the inventory on hand at an FSL, an emergency order is immediately placed on the ESL for an amount equal to the customer's order size. If the ESL does not have enough inventory on hand, the excess quantitiy becomes a backorder at the ESL. Upon receipt of a resupply request placed by a FSL for a given amount of stock, the ESL in turn places a replenishment order for the same amount on the RSL. Thus a customer's order may be satisfied by one of two different types of replenishment orders depending on whether or not the on hand inventory at the FSL is sufficient to satisfy the customer's demand. Ad-

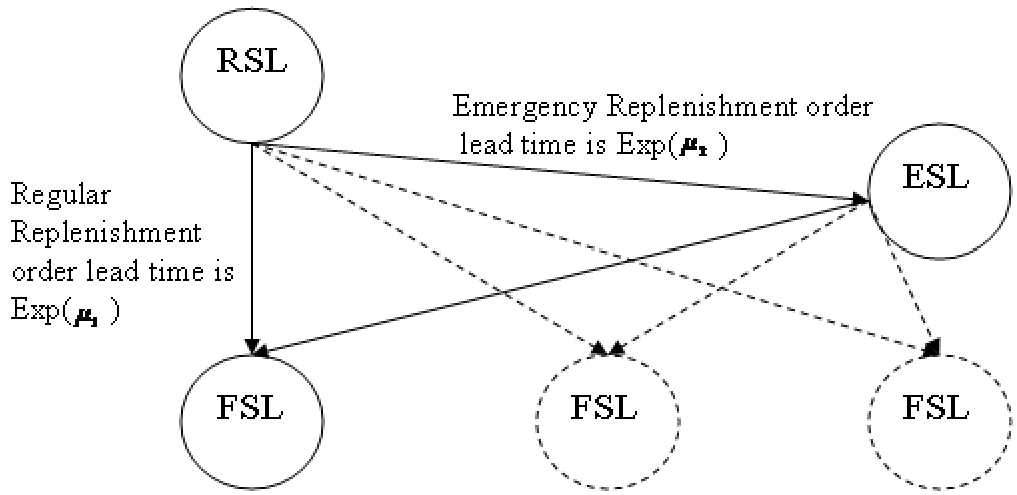

Figure 4.1: A System with Emergency Resupply

ditionally, we assume that the lead times from the RSL to the FSLs and the ESL are exponentially distributed. Figure 4.1 depicts the resupply system we have described.

To determine the optimal order-up-to levels at the FSLs and the ESL, we require the stationary distributions of the number of units in regular and emergency resupply. Chen.*et al.* (2010a,2010b) analyze the same system and propose both an exact and an approximate method for determining these distributions. Observe that an emergency order can be treated as a lost sale at the FSL since the FSL replenishment process from the RSL is equal to the amount of stock withdrawn for the FSL inventory, which corresponds to customer's demand can be satisfied totally from FSL inventory. Chen.*et al.* (2010a) analyze such a lost sales system and derive the exact steady state distribution of the number of units in regular resupply of a field service location that employs an (s-1,s) inventory policy. Furthermore, Chen.*et.al.* (2010b) use a zero-truncated negative binomial distribution with an atom at zero to approximate the steady state distribution of the number of emergency ordered units outstanding in a system consisting of both FSLs and an ESL.

Our goal in this paper is to develop optimization algorithms for setting stock levels in the emergency resupply network that we have described. The exact and an approximate stationary distribution of the number of outstanding ordered units at the FSLs and ESL are used to construct the expected cost function corresponding to the service network system. The costs considered in the optimization model are emergency order penalty costs for each order placed by an FSL on the ESL, and backorder penalty costs charged at the ESL for each backordered unit per unit of time. Since each time an order arrives at the FSLs, a regular or an emergency replenishment order is placed at the RSL therefore the inventory in the FSLs and ESL system is kept at a constant level. We use a budget constraint instead of a holding cost to capture the cost associated with the inventory handled in the whole system. We show that the emergency order cost function at the FSL is non-increasing and observe its convexity in its targeted inventory level. The backorder costs charged at the ESL depend on the steady state distribution of the number of units in emergency resupply which in turn, is influenced by the target inventory levels at the FSLs. These backorder costs are non-increasing in the budget allocated to the ESL, for given FSL stock levels. Therefore, in our algorithm, we set the inventory levels at the FSLs first and then allocate all the remaining budget to the ESL. Using these ideas, we develop a simple bisection search method to set the desired stock levels.

The reminder of the paper is organized as follows. In section 2 we briefly review the results discussed in Chen. *et al.* (2010b) and extend them to the case where there are multiple part types. In section 3, we formulate the objective function and in the following section 4, we build mathematical models and explain our optimization approach. In section 5, we develop algorithms to search for the optimal stock levels for a single part type. Experimental results are shown in section 6. Further discussion on the optimization algorithm for multi-part types is in section 7. Concluding comments are found in section 8.

4.1.2 Literature Review

Sherbrooke (1966) builds a mathematical model for the inventory control of recoverable or repairable items in a base-depot supply system. The model is well known as METRIC (Multi-Echelon Technique for Recoverable Item Control) and is extended and improved by amount of subsequent papers such as Graves (1985) and Sherbrooke (1986). The METRIC model uses negative binomial distribution to approximate the stationary distribution of the number of units in resupply, which largely simplifies the computational complexity. Two-phase marginal analysis algorithm is used to determine the depot and base stock levels, which optimize the objective function under investment constraints. Alternatively, Fox and Landi (1970) propose a Lagrangian multiplier method for solving the one-constraint optimization problems as in METRIC by using one pass method to search for the suitable multipliers. Muckstadt (1978) suggests a simple approximation for the optimization problem and develops an easier algorithm to determine the stock levels compared with the previous two methods. More details are discussed and summarized in Muckstadt (2005,2010).

One of the basic assumptions in these multi-echelon resupply networks is that each location has a single source of resupply. However, there are numerous examples in practice that locations share inventories among themselves (*lateral transshipment*) or obtain emergency orders from alternative sources (emergency stocking locations). Early studies consider emergency lateral transshipment include Gross (1963), Das (1975), Hoadly and Heyman (1977), Karmarkar and Patel (1977), Cohen *et al.* (1986), Dada (1985), Bowman(1986), and Slay (1986). Using the pooling idea of Cohen *et al.*, Lee (1987) extends the METRIC model so that the out-of-stock bases could get emergency lateral transshipment from other identical bases with inventories in the same group. If all bases in the group have zero inventory, the current demand is sent to the depot. Approximations for the system performance measures, such as backorder level and the number of emergency lateral transshipment, are derived and used to optimize the stocking levels with two-phase method. Axsäter (1990) applies alternative method to model the demand at the bases allowing non-identical bases and compares the results with Lee's when the bases are identical.

In contrast to the military base-depot model as METRIC, Grahovac and Chakravarty (2001) present a commercial supply chains allowing emergency orders and lateral transshipment. When the inventory at the retailers is below some point K, they place emergency orders from their upstream distribution center. An emergency transshipment is requested only when the distribution center runs out of stock and at least one retailer has more than K inventory on hand. With the guaranteed and expedited shipment delivery service, this model could prevent unnecessary lateral transshipment and complicate transaction.

There are more papers examining the effect of employing decision rules for making lateral transshipment, such as Dada (1992), Sherbrooke (1992), Evers (1997,1999), Alfredsson and Verrijdt (1999), Kukreja *et al.* (2001), Muckstadt

(2005), Vidgren (2005), Axsäter (2006) and Vliegen (2009). Another type of papers presents methods for optimizing the decisions concerning lateral transshipment. Examples are Das (1975), Robinson (1990), Tagaras and Cohen (1992), Archibald *et al.* (1997), Rudi *et al.* (2001), Minner *et al.* (2003), Wong *et al.* (2006), Olsson (2009), Kranenburg and van Houtum (2009), Wijk *et al.* (2009) and Reijnen *et al.*(2009). Paterson *et al.* (2009) provide an up-to-date review of the inventory models with lateral transshipment.

Different from the resupply networks considering both lateral transshipment and emergency orders from upstream or external supplier, the system considered in this paper contains special stocking location, the ESL, which is dedicated to satisfying emergency orders. Once the FSL is out of stock, it can only fulfill the arriving customer by placing emergency orders from the ESL. No lateral transshipment is allowed among the FSLs at any time. It could also be interpreted as that the inventory shared among the FSLs is stocked at the ESL and is consumed only when one FSL incurs shortage. The systems allowing emergency orders have also been widely studied such as the papers by Rosenshine and Obee (1976), Whittemore and Saunders (1977), Blumenfeld *et al.* (1985), Moinzadeh and Schmidt (1991), Johansen and Thorstenson (1998), Tagaras and Vlachos (2001), Chiang (2002), Axsäter (2007), etc. Refer to Chen *et al.* (2010) for an overview of these papers. Most of the literature focuses on the optimal inventory policy and the replenishment orders modeling with a single-echelon. The emergency orders are placed either from the same source of the regular replenishment or external supplier with infinite inventory. In this paper, the inventory at the ESL is limited and backorder costs at the ESL is included in the objective function. We develop optimization algorithms to determine the optimal stocking levels for this multi-item two-echelon network and investigate the advantage of the ESL under different variance-to-mean-ratio demand scenarios.

4.2 Model Review

To analyze the system we have described, we make a number of simplifying assumptions. As we have stated, we assume that the RSL has an infinite stock of multiple types of items on hand and the lead time from the ESL to the FSL is negligible. Assuming that the demand of the different types of items arrive independently, we are able to construct separable cost and customer service measures that depend on the steady state distributions of the number of units in regular replenishment from the RSL to the FSL (i.e. in regular resupply) and the number of units in regular replenishment from the RSL to the ESL (i.e. in emergency resupply). We use the complete fill assumption at the FSLs instead of partial fill to maintain consistency among these papers. We could easily extend our results to the partial fill case.

Let $\lambda^{(n,i)}$ denote the rate of customer arrivals at the nth FSL ($n = 1, 2, \ldots, N$) for type i part ($i = 1, 2, \ldots, I$). Let $X^{(n,i)}$ denote the size of any customer's order, a positive, integer-valued, random variable. Let $p_k^{(n,i)} \equiv P\left\{X^{(n,i)} = k\right\}$ and let $\overline{P}_k^{(n,i)} \equiv P\left\{X^{(n,i)} > k\right\}$ for all $k = 0, 1, 2, \ldots$. Since the arrival processes are stuttering Poisson processes, $X^{(n,i)}$ is geometrically distributed, that is, $p_k^{(n,i)} = (1 - p^{(n,i)})^{k-1} p^{(n,i)}$, where $p^{(n,i)} = p_1^{(n,i)}$ for $k = 1, 2, \ldots$ and $p_0^{(n,i)} = 0$. We can easily generalize our results to allow for zero-sized orders.

Let $I_t^{(n,i)}$ denote the inventory of item i on hand at the nth FSL if $n > 0$ or the ESL if $n = 0$, at time t, $t \geq 0$, a non-negative integer-valued random variable. Recall that the system is managed according to an $(S-1, S)$ policy. Suppose a customer arrives the nth FSL at time t with demand for part type i, denoted by $X_t^{(n,i)}$. If $X_t^{(n,i)} \leq I_t^{(n,i)}$ the demand is satisfied by the inventory at the nth FSL and triggers a regular replenishment order from the RSL to the nth FSL with size $X_t^{(n,i)}$; otherwise, it is filled by the ESL and the ESL places a replenishment order from the RSL of size $X_t^{(n,i)}$. Our decision variables are the stock up to levels, denoted by $S^{(n,i)}$, of part type i at the nth FSL ($n > 0$) or the ESL ($n = 0$).

From Chen. *et al.* (2009a), we have the following result:

Let $f_{NB}(\cdot; m, p)$ denote the negative binomial probability distribution with parameters m and p :

$$f_{NB}(x; m, p) \equiv \binom{m + x - 1}{x} p^m (1 - p)^x \text{ for} x = 0, 1, 2, \ldots.$$

Proposition 14 *For a lost sales system with stuttering Poisson demand with complete fills and targeted inventory level $S^{(n,i)}$, the stationary distribution of the number of units of item i at the nth FSL on order, denoted by $\pi_{s|S^{(n,i)}}$ ($s = 0, 1, \ldots, S^{(n,i)}$), is given by:*

$$\pi_{s|S^{(n,i)}} = \frac{\sum_{m=0}^{s} (\frac{\lambda^{(n,i)}}{\mu})^m \frac{f_{NB}(s-m;m,p^{(n,i)})}{m!}}{G(S^{(n,i)})},$$

where $S^{(n,i)}$ is the stock up to level, $G(S) = \sum_{s=0}^{S} \sum_{m=0}^{s} \frac{(\frac{\lambda^{(n,i)}}{\mu})^m}{m!} f_{\text{NB}}(s - m; m, p^{(n,i)})$, and $f_{\text{NB}}(s - m; 0, p) = 1\{s = 0\}$ when $m = 0$. i.e. the truncated compound Poisson distribution.

The notation $\pi_{s|S^{(n,i)}}$ emphasizes its dependence on the value of $S^{(n,i)}$.

Chen. *et al.* (2009b) show how to approximate the steady state distribution of the number of units of type i in emergency resupply by a zero-truncated negative binomial distribution with an atom at zero given the stock levels of each item at the FSLs: $\vec{S}^{(i)} = (S^{(1,i)}, S^{(2,i)}, \ldots, S^{(N,i)})$. Denoting this approximate distribution function by $f_z(\vec{S}^{(i)})$ for $z = 0, 1, \ldots$, we have

$$f_z(\vec{S}^{(i)}) = \begin{cases} f_0, & \text{if} z = 0, \\ (1 - f_0)[\frac{1}{1-(p_E)^{r_E}}\binom{z+r_E-1}{z}(p_E)^{r_E}(1-p_E)^z], & \text{if} z > 0, \\ 0, & \text{otherwise.} \end{cases}$$

where f_0, r_E and p_E depend on $\vec{S}^{(i)}$.

Next, we will use the steady state distributions $\pi_{s|S^{(n,i)}}$ and $f_z(\vec{S}^{(i)})$ to construct the cost function.

4.3 Objective Function Formulation

As we have mentioned, our objective function consists of the emergency order penalty cost and a backorders penalty cost at the ESL. Define $\vec{S}_T^{(i)} = (\vec{S}^{(i)}, S^{(0,i)})$ and $\vec{S}_T = (\vec{S}_T^{(1)}, \vec{S}_T^{(2)}, \ldots, \vec{S}_T^{(I)})$.

Let $c_E^{(i)}$ denote the emergency order cost per backordering incident per order of part type i at the FSLs. The expected emergency order penalty at the nth FSL equals

$$\begin{aligned} c_{FSL}^{(n,i)}(S^{(n,i)}) &\equiv c_E^{(i)}\lambda^{(n,i)}\{\textstyle\sum_{s=0}^{S^{(n,i)}} P[X^{(n,i)} > S^{(n,i)} - s]\pi_{s|S^{(n,i)}}\} \\ &= c_E^{(i)}\lambda^{(n,i)}\textstyle\sum_{s=0}^{S^{(n,i)}}(1 - p^{(n,i)})^{S^{(n,i)}-s}\pi_{s|S^{(n,i)}}, \end{aligned} \tag{4.1}$$

and the expected emergency order penalty at FSLs for part type i is given by

$$C_{FSL}^{(i)}(\vec{S}^{(i)}) = \sum_{n=1}^{N} c_{FSL}^{(n,i)}(S^{(n,i)}) \tag{4.2}$$

which does not depend on $S^{(0,i)}$. The total expected emergency order penalty at the FSLs is $C_{FSL}(\vec{S}_T) \equiv \sum_{i=1}^{I} C_{FSL}^{(i)}(\vec{S}_T^{(i)}) = \sum_{i=1}^{I}\sum_{n=1}^{N} c_{FSL}^{(n,i)}(\vec{S}_T^{(n,i)})$.

Let $c_B^{(i)}$ denote the backorder cost per unit time per unit of part type i backordered at the ESL. The expected backorder costs for part type i given $\vec{S}_T^{(i)}$ is denoted by

$$C_{ESL}^{(i)}(\vec{S}_T^{(i)}) \equiv c_B^{(i)} E[(z - S^{(0,i)})|\vec{S}^{(i)}] = c_B^{(i)}[\sum_{z=S^{(0,i)}}^{\infty} (z - S^{(0,i)}) f_z(\vec{S}^{(i)})].$$

The total expected backorder cost is $C_{ESL}(\vec{S}_T) \equiv \sum_{i=1}^{I} C_B^{(i)}(\vec{S}_T^{(i)})$.

Therefore the total expected cost associated with part type i is

$$\begin{aligned} C^{(i)}(\vec{S}_T^{(i)}) &\equiv C_{FSL}^{(i)}(\vec{S}^{(i)}) + C_{ESL}^{(i)}(\vec{S}_T^{(i)}) \\ &= c_E^{(i)} \textstyle\sum_{n=1}^{N} \lambda^{(n,i)} \sum_{s=0}^{S^{(n,i)}} (1 - p^{(n,i)})^{S^{(n,i)}-s} \pi_{s|S^{(n,i)}} + c_B^{(i)} \sum_{z=S^{(0,i)}}^{\infty} (z - S^{(0,i)}) f_z(\vec{S}^{(i)}), \end{aligned}$$

and the total cost for the system with multiple part types is

$$C(\vec{S}_T) = \sum_{i=1}^{I} C^{(i)}(\vec{S}_T^{(i)}) = \sum_{i=1}^{I} C_{FSL}^{(i)}(\vec{S}^{(i)}) + C_{ESL}^{(i)}(\vec{S}^{(i)}, S^{(0,i)}).$$

Proposition 15 *For any (n, i), the function $c_{FSL}^{(n,i)}(S^{(n,i)})$ is non-increasing in $S^{(n,i)}$.*

Proof: Without loss of generality, we drop the superscript (n, i). The value $c_{FSL}(S)$ is the expected emergency order cost given the target stock level S. For the FSL with target stock level $S + 1$, if the inventory policy of the FSL is changed and it is not allowed to use the last unit on hand at the FSL (Scenario 1), the resupply process is exactly the same as that of the FSL employing (S-1,S) inventory policy with target stock level S (Scenario 2). For any sample path of the arrival process, the emergency order cost is the same for both scenarios. Now given any sample path of the arrival process, if the spare unit is consumed at any point in time and never resupplied, the corresponding emergency order cost is non-increased. Furthermore, if the spare unit is resupplied later and could be consumed again, the corresponding emergency order cost for the same sample path should be not larger than the no resupply case. Therefore, given sample path of the arrival process, the emergency order cost for the FSL employing (S-1,S) inventory policy with target stock level $S + 1$ is smaller than the cost for the Scenario 1 FSL in Scenario 1, or the Scenario 2 FSL, which employs (S-1,S) inventory policy with target stock level S. Thus, the expected emergency order cost $c_{FSL}(S)$ is non-increasing in S. ∎

In addition, empirical investigation of $c_{FSL}(S)$ suggests that the function is convex in S over a wide range of parameters. Consequently, we use algorithms that exploit this apparent convexity. Should a case emerge in which this function is found to be non-convex, we recommend using the largest convex minorant of the true function.

Proposition 16 *For any i, the function $C^{(i)}_{ESL}(\vec{S}^{(i)}, S^{(0,i)})$ is non-increasing and convex in $S^{(0,i)}$ when $\vec{S}^{(i)}$ is fixed.*

The result could be easily proved by using first order differences.

4.4 Mathematical Modeling

As seen in proposition 15 and 16, it seems that the system should set the inventory levels at the FSLs and ESL as high as possible to minimize the associated emergency order penalty costs and backorder penalty costs. However, in real life situations, limits often exist on the system investment in inventory over all part types due to the holding costs and capital limits. Given the investment limits, we then have to balance the stock levels of different part types to minimize the overall expected costs. Let B denote the fixed inventory investment budget. Since the cost functions are non-increasing, the optimal targeted inventory stock levels should sum up to B. Let Z^+ be the state space of nonnegative integers. The mathematical model of the whole system is as follows:

$$\begin{aligned} \min_{\vec{S}_T} \quad & C(\vec{S}_T) \\ \text{s.t.} \quad & \textstyle\sum_{i=1}^{I} \sum_{n=0}^{N} S^{(n,i)} = B, \\ & S^{(n,i)} \in Z^+ \text{for} n = 0, 1, \ldots, N; i = 0, 1, \ldots, I. \end{aligned}$$

Our approach is to minimize the cost of each product $C^{(i)}(\vec{S}_T^{(i)})$ given a specified budget $B^{(i)}$ for part type i first, and then to minimize the overall cost $C(\vec{S}_T)$ by varying $\vec{B} = (B^{(1)}, \ldots, B^{(I)})$ where $\sum_{i=1}^{I} B^{(i)} = B$. Therefore, the model is changed into:

$$\begin{aligned} \min_{\vec{B}} \quad & \textstyle\sum_{i=1}^{I} \min_{\{\sum_{n=0}^{N} S^{(n,i)} = B^{(i)}\}} C^{(i)}(\vec{S}_T^{(i)}) \\ \text{s.t.} \quad & \textstyle\sum_{i=1}^{I} B^{(i)} = B, \\ & B^{(i)} \in Z^+ \text{for} i = 0, 1, \ldots, I. \end{aligned} \tag{4.3}$$

Hence, for each part type i, we have the subproblem (i):

$$\begin{aligned} \min_{\vec{S}_T^{(i)}} \quad & C^{(i)}(\vec{S}_T^{(i)}) \\ \text{s.t.} \quad & \textstyle\sum_{n=0}^{N} S^{(n,i)} = B^{(i)}, \\ & S^{(n,i)} \in Z^+ \text{for} n = 0, 1, \ldots, N, \end{aligned}$$

which is the same as

$$\begin{aligned} G^*(B^{(i)}) \equiv \min_{\vec{S}^{(i)}} \quad & C_{FSL}(\vec{S}^{(i)}) + C_{ESL}(\vec{S}_T^{(i)}) \\ \text{s.t.} \quad & \textstyle\sum_{n=1}^{N} S^{(n,i)} + S^{(0,i)} = B^{(i)}, \\ & S^{(n,i)} \in Z^+ \text{for} n = 0, 1, \ldots, N. \end{aligned} \tag{4.4}$$

Define the optimal minimizer as $\vec{S}^*(B^{(i)})$. Now, our problem (4.3) is equivalent to

$$\begin{aligned} \min_{\vec{B}} \quad & \textstyle\sum_{i=1}^{I} G^*(B^{(i)}) \\ \text{s.t.} \quad & \textstyle\sum_{i=1}^{I} B^{(i)} = B, \\ & B^{(i)} \in Z^+ \text{for} i = 0, 1, \ldots, I. \end{aligned} \tag{4.5}$$

Our goal is to construct the maximal convex minorant of function $G^*(B^{(i)})$ for each part type i and then to use marginal analysis to solve the resulting optimization problem (4.5). After obtaining the optimal allocation of the budget among the I part types $\vec{B}^* = (B^{*(1)}, \ldots, B^{*(I)})$, the optimal stock levels on the FSLs are the corresponding values of the $\vec{S}^*(B^{*(i)})$ from (4.4) for the associated budgets $B^{*(i)}$.

4.4.1 Optimize Order-up-to-Levels at Different Locations for a Single Item

In this section, we demonstrate a method to approximate the function $G^*(B^{(i)})$ for a given part type i (problem(4.4)). To simplify notation, we drop the superscript (i) and use S_n instead of $S^{(n,i)}$. Let $\vec{S} = (S_1, \ldots, S_N)$ represent the stock levels for the FSLs. The optimization problem we wish to solve, (4.4), is rewritten as

$$\begin{aligned} G^*(B) \equiv \min_{\vec{S}} \quad & C_{FSL}(\vec{S}) + C_{ESL}(\vec{S}_T) \\ \text{s.t.} \quad & \textstyle\sum_{n=1}^{N} S_n + S_0 = B, \\ & S_n \in Z^+ \text{for} n = 0, \ldots, N. \end{aligned} \tag{4.6}$$

Let B_F denote the total inventory at the FSLs. The the optimal stock level at the ESL should be $S_0 = B - B_F$. That is, all of the remaining budget should be allocated to the ESL due to the non-increasing nature of the backorder cost function. The expected cost given B_F and B is denoted as $J^*(B_F, B)$:

$$\begin{aligned} J^*(B_F, B) \equiv \min_{\vec{S}} \quad & C_{FSL}(\vec{S}) + C_{ESL}(\vec{S}, B - B_F) \\ \text{s.t.} \quad & \textstyle\sum_{n=1}^{N} S_n = B_F \\ & S_n \in Z^+ \text{for} n = 1, \ldots, N. \end{aligned}$$

It follows that $G^*(B) = \min_{B_F \leq B} J^*(B_F, B)$.

Due to the time consuming step of matrix inversion used to analyze the behavior of the ordered units at the ESL (Chen, *et al.* 2010b), it takes a much longer time to compute $C_{ESL}(\vec{S}, B - B_F)$ than to compute $C_{FSL}(\vec{S})$ given any $\vec{S}$. To mitigate this problem, we define the following alternative optimization problem focusing on the FSLs:

$$\begin{aligned} H_F^*(B_F) \equiv \min_{\vec{S}} \quad & C_{FSL}(\vec{S}) \\ \text{s.t.} \quad & \textstyle\sum_{n=1}^{N} S_n = B_F \\ & S_n \in Z^+ \text{for} n = 1, \ldots, N. \end{aligned} \tag{4.7}$$

Denote its optimal solution as $\vec{S}^*(B_F)$. Instead of solving problem (4.6), we then use

$$\tilde{G}^*(B) \equiv \min_{B_F \leq B} \tilde{J}^*(B_F, B) = \min_{B_F \leq B} H_F^*(B_F) + C_{ESL}(\vec{S}^*(B_F), B - B_F). \tag{4.8}$$

as an approximation to $J^*(B_F, B)$ given any B and $B_F \leq B$.

Recall from (4.1) and (4.2) that

$$\begin{aligned} C_{FSL}(\vec{S}) &= \textstyle\sum_{n=1}^{N} c_{FSL}^{(n)}(S_n) \\ &= c_E \textstyle\sum_{n=1}^{N} \sum_{s=0}^{S_n} [\lambda^{(n)}(1 - p^{(n)})^{S_n - s}] \pi_{s|S_n}, \end{aligned} \tag{4.9}$$

where $c_{FSL}^{(n)}(S_n)$ is non-increasing (proposition 15). As noted, numerical experiments suggest $c_{FSL}^{(n)}(S_n)$ convex in S_n. Therefore, we use marginal analysis to solve the optimization problem (4.7) given any B_F. Denote $B_F^*(B)$ as the optimal solution of problem (4.7), as found by marginal analysis.

Next we use a line search method such as bisection or golden section search to solve $\tilde{G}^*(B) = \min_{B_F \le B} \tilde{J}^*(B_F, B)$ as an approximation of $G^*(B)$ given B. Denote the corresponding optimized inventory levels which depend on B by

$$\vec{S}_T(B_F^*(B)) = (\vec{S}(B_F^*(B)), B - B_F^*(B)).$$

Hence we find $\tilde{G}^*(B)$ through a combination of marginal analysis over S_n nested within a line search over B_F.

It is conceivable that the optimal solution to problem (4.6) does not simultaneously optimize problem (4.8), i.e. $G^*(B) \neq \tilde{G}^*(B)$. To explore the difference, we compare the solutions to (4.8) with the best solutions to (4.6) found using a more comprehensive search algorithm. The particular search algorithm used as a benchmark is the Particle Swarm Pattern Search method (PSWARM), developed by Vaz and Vicente(2007,2009) for solving minimization problem subject to simple bounds (linear constraints) without the use of derivatives. We find that by solving problem (4.8), we are able to obtain solutions close to the benchmark results.

4.5 Heuristic Algorithms Given Inventory Investment for Single Item

In this section, we describe the methods sketched in the previous section.

- (H1) "**Nested Search**": Experimentation strongly suggests that $\tilde{J}^*(B_F, B)$ is unimodal in B_F. Consequently, we use a bisection search for B_F minimizing $\tilde{J}^*(B_F, B)$ instead of computing $C_{ESL}(\vec{S}, B - B_F)$ exhaustively for all possible values of B_F. This algorithm has two steps: Step one, use marginal analysis to determine $H_F^*(B_F)$ and $\vec{S}^*(B_F)$ for $B_F = 0, \ldots, B$; Step two, use bisection search to determine the optimal $B_F^*(B)$ which minimizes $\tilde{J}^*(B_F, B) = H_F^*(B_F) + C_{ESL}(\vec{S}, B - B_F)$ where B is known and fixed.

Step One: Marginal Analysis

1. Start with $B_F = 0$ and $\vec{S} = (0, \ldots, 0)$. Define $g(B_F) = \vec{S}$.
2. For all n, compute $\Delta c_{FSL}^{(n)}(S_n) = c_{FSL}^{(n)}(S_n + 1) - c_{FSL}^{(n)}(S_n)$ and let $n^* = argmin_n \Delta c_{FSL}^{(n)}(S_n)$.
3. Update $S_{n*} = S_{n^*} + 1$, $B_F = B_F + 1$ and $g(B_F) = \vec{S}$. If $B_F < B$ then go back to 2. Otherwise, go to Step Two.

Step Two: Bisection Search

1. Choose initial interval [a,b] over which the minimum of $g(B_F) = C_{FSL}(\vec{S}^*(B_F)) + C_{ESL}(\vec{S}^*(B_F), B - B_F)$ is to be found. For example, $a = 0$ and $b = B$. Choose separation constant ε and stopping tolerance θ.
2. If $b - a < \theta$, then go to step 4. Otherwise, let $a' = (a + b)/2 - \varepsilon$ and $b' = (a + b)/2 + \varepsilon$. Round a', b' to their nearest integers.
3. If $g(a') < g(b')$, then update $b = b'$. Otherwise, update $a = a'$. Then go back to 2.
4. Identify $B_F^*(B) = argmin_{\{x=a,a',b,b'\}} g(x)$ and $\tilde{G}^*(B) = g(B_F^*(B))$.

Validate the method of approximating $G^*(B)$ by $\tilde{G}^*(B)$, we use PSWARM to search over the state space

$$\mathfrak{S}(B) = \{\vec{S}_T \in \{Z^+, \ldots, Z^+\}_{1 \times N+1} : \sum_{n=1}^{N} S_n + S_0 = B\}.$$

Refer to Vaz and Vicente(2007,2009) for more details of PSWARM. A high level view of the algorithm is as follows:

- (H2) "**PSWARM Search**":
  1. Pick an initial point $\vec{S}_T$ in $\mathfrak{S}(B)$.
  2. Use PSWARM algorithm to search over $\mathfrak{S}(B)$ for the optimal solution $\vec{S}^*_T(B) = (\vec{S}^*(B), S^*_0(B))$ which minimizes $C_{FSL}(\vec{S}) + C_{ESL}(\vec{S}_T)$.
  3. Return $G^*(B) = C_{FSL}(\vec{S}^*(B)) + C_{ESL}(\vec{S}^*_T(B))$.

4.6 Experimental Results

We conduct experiments with five FSLs, $N = 5$, for a single part type in the system. We fix the lead time, $\tau_F = \tau_E$, equal to 1. The backorders cost parameter is $C_B = 20$, and the emergency order penalty, C_E, is one of the values $0.25, 1, 5$ or 10.

To describe the arrival processes, let $W_n(t)$ denote the cumulative unit arrivals during time t for the nth FSL. Let μ_n denote the arrival process mean rate, which is equal to $\frac{E(W_n(t))}{t} = \frac{\lambda_n}{p_n}$. Let σ_n^2 denote the arrival process variance rate, which is equal to $\frac{Var(W_n(t))}{t} = \lambda_n(\frac{1-p_n}{p_n^2} + \frac{1}{p_n^2})$. Let ρ_n denote the arrival process variance-to-mean ratio (VTMR), which is equal to

$$\frac{Var(W_n(t))/t}{E(W_n(t))/t} = \frac{\sigma_n^2}{\mu_n} = \frac{2 - p_n}{p_n}.$$

We study cases both of identical FSLs and non-identical FSLs. After considering these cases, we explore the shape of the cost function $\tilde{G}^*(B)$ and recommend using a convex minorant of this function for solving multiple item problems.

4.6.1 Identical Independent FSLs

First, we conduct an experimental study assuming that the demand distribution is identical for all FSLs. Therefore μ_n and ρ_n are common for $n = 1, \ldots, 5$. We choose μ_n to be equal to one of 1, 5, or 10 and ρ_n to be equal to one of 1.01, 2, or 5. The investment budget B is fixed and equal to 80. For this identical independent FSLs case, our heuristic algorithm (H1) solving problem (4.8) and the benchmark PSWARM algorithm (H2) solving problem (4.6) all lead to the same optimal solutions, i.e. $G^*(B) = \tilde{G}^*(B)$.

Table 4.1 shows the optimal solutions of stock levels, $\vec{S}_T = [\vec{S}, S_0]$, and the optimized cost, $G^*(B) = \tilde{G}^*(B)$. From the experimental results, we see that the optimal stock levels of the FSLs are not necessary identical but the stock level differences are at most one. For each combination of (μ_n, ρ_n), we observe that as the emergency order penalty C_E increases, the optimal stock levels at the FSLs increases. This means less inventory is kept at the ESL. Furthermore, the increase in C_E also causes the total expected cost $G^*(B)$ to increase.

When the demand mean, μ_n is held constant but the variance-to- mean ratio ρ_n increases, the total expected cost $G^*(B)$ increases and more inventory is stocked at the ESL.

When the variance-to-mean ratio, ρ_n, is held constant but the mean, μ_n, increases, the total expected cost $G^*(B)$ also increases and more inventory is

Table 4.1: The Optimal Inventory Levels and Expected Total Cost for the Identical Independent FSLs case when $B = 80$:

		$\rho_n = \frac{\sigma_n^2}{\mu_n} = 1.01$			$\rho_n = \frac{\sigma_n^2}{\mu_n} = 2$			$\rho_n = \frac{\sigma_n^2}{\mu_n} = 5$		
μ_n	C_E	$\vec{S}^*(B)$	S_0^*	$G^*(B)$	$\vec{S}^*(B)$	S_0^*	$G^*(B)$	$\vec{S}^*(B)$	S_0^*	$G^*(B)$
1	0.25	[16 16 16 16 16]	0	0	[13 13 13 13 13]	15	0.000147	[10 10 10 10 10]	30	0.0207
	1	[16 16 16 16 16]	0	0	[14 14 13 13 13]	13	0.000416	[11 11 11 10 10]	27	0.0692
	5	[16 16 16 16 16]	0	0	[14 14 14 14 14]	10	0.00119	[12 12 11 11 11]	23	0.272
	10	[16 16 16 16 16]	0	0	[15 15 14 14 14]	8	0.00197	[12 12 12 12 12]	20	0.482
5	0.25	[16 15 15 15 15]	4	0.00111	[14 14 13 13 13]	13	0.0803	[6 6 6 5 5]	52	1.14
	1	[16 16 15 15 15]	3	0.00376	[14 14 14 14 14]	10	0.252	[8 8 8 8 8]	40	2.76
	5	[16 16 16 15 15]	2	0.0149	[15 15 14 14 14]	8	0.990	[11 11 10 10 10]	28	8.40
	10	[16 16 16 16 15]	1	0.0271	[15 15 15 15 14]	6	1.78	[12 11 11 11 11]	24	14.0
10	0.25	[14 13 13 13 13]	14	1.29	[8 8 7 7 7]	43	4.76	[0 0 0 0 0]	80	11.8
	1	[14 14 14 14 14]	10	3.55	[11 10 10 10 10]	29	11.4	[1 1 1 1 1]	75	24.0
	5	[15 15 15 15 15]	5	12.1	[13 13 13 13 12]	16	32.7	[6 6 6 6 5]	51	64.5
	10	[16 15 15 15 15]	4	21.2	[14 14 13 13 13]	13	52.8	[8 8 8 8 8]	40	98.6

stocked at the ESL as well.

4.6.2 Non-Identical Independent FSLs

We conduct another experiment study assuming that the five FSLs have different arrival process as shown in Table 4.2. The parameters are chosen to represent the different scenarios that might be encountered in real life: low mean demand with medium variance (LM), low mean demand with high variance (LH), medium mean demand with medium variance (MM), high mean demand with low variance (HL), and high mean demand with medium variance (HM).

In Table 4.3, we show the optimal solutions of stock levels $\vec{S}_T = [\vec{S}, S_0]$ and the optimal cost $\tilde{G}^*(B)$ solved by algorithm (H1) and $G^*(B)$ solved by algorithm (H2). The investment budget B is equal to either 30, 50 or 80.

For this case of non-identical FSLs, the optimal solution of problem (4.6)

Table 4.2: Parameters of the Non-Identical Demand Arrival Processes at the FSLs :

n	1	2	3	4	5
Demand Type	LM	LH	MM	HL	HM
μ_n	1	1	5	10	10
σ_n^2	2	5	10	10.1	20
$\rho_n = \frac{\sigma_n^2}{\mu_n}$	2	5	2	1.01	2
λ_n	0.667	0.333	3.33	9.95	6.67
p_n	0.667	0.333	0.667	0.995	0.667

Table 4.3: The Optimal Inventory Levels and Expected Total Cost for the Non-Identical Independent FSLs case:

		Algorithms (H1)			Algorithm (H2)			Relative Error
B	C_E	$\vec{S}^*(B)$	S_0^*	$\tilde{G}^*(B)$	$\vec{S}^*(B)$	S_0^*	$G^*(B)$	$\frac{\tilde{G}^*(B)-G^*(B)}{G^*(B)}$
30	0.25	[0 0 0 0 0]	30	36.2	[0 0 0 0 0]	30	36.2	0 %
	1	[0 0 0 4 0]	26	50.9	[0 0 0 4 0]	26	50.9	0 %
	5	[0 0 1 10 5]	14	100	[0 0 1 9 6]	14	99.6	0.564 %
	10	[0 0 3 11 7]	9	139.9	[0 0 3 10 8]	9	139.6	0.196 %
50	0.25	[0 0 4 12 9]	25	2.10	[0 0 3 12 12]	23	2.05	2.80 %
	1	[0 0 6 14 13]	17	4.84	[0 0 6 13 14]	17	4.82	0.404 %
	5	[1 0 8 15 15]	11	14.0	[1 0 8 15 15]	11	14.0	0 %
	10	[2 0 8 16 16]	8	23.1	[2 0 9 15 16]	8	22.9	0.908 %
80	0.25	[4 4 13 19 21]	19	0.111	[4 3 13 19 22]	19	0.106	4.36 %
	1	[4 4 13 19 22]	18	0.345	[5 3 13 20 23]	16	0.332	3.77 %
	5	[5 6 14 20 23]	12	1.22	[5 5 14 20 23]	13	1.21	0.958 %
	10	[5 7 15 20 23]	10	2.09	[5 7 14 20 23]	11	2.09	0.188 %

is different from but close to that of (4.8) for most cases. The relative error, $\frac{\tilde{G}^*(B)-G^*(B)}{G^*(B)}$ is bounded by 5% and less than 1% for most cases. The approximation is especially good when the budget B is small. Combined with the results of the identical independent FSLs, it suggests that $\tilde{G}^*(B)$ is a reasonable approximation of $G^*(B)$ and therefore the bisection search algorithm (H2) is recommended to save computing time.

When the investment budget B is low and equal to 30, no inventory is kept at the FSLs with low demand rate. Instead, inventories are concentrated at the FSLs with high mean and at the ESL. The FSL with high mean and medium variance receives less stock allocation than the FSL with high mean and low variance.

When the investment budget B increases to 50, the optimal stock levels for low mean demand are still very small. In particular, the optimized solution still chooses to keep zero inventory at the FSL which has a low mean and high variance. In addition, with increased budget, less inventory is stocked at the ESL.

When the investment budget B increases to 80, the optimal stock levels increases significantly for all the FSLs. The FSL with a high mean demand stocks more if its variance is higher. This suggests that if the investment budget is large enough, the optimal stock levels are positively correlated with both the mean and variance of the demand. As a general rule, our intuition is that when the investment budget B is large enough, all the inventory should be kept at the FSLs to prevent out-of-stock/emergency orders and no inventory is needed at the ESL. At the same time, the FSLs with higher variance need to stock more than FSLs with the same mean level but low variance.

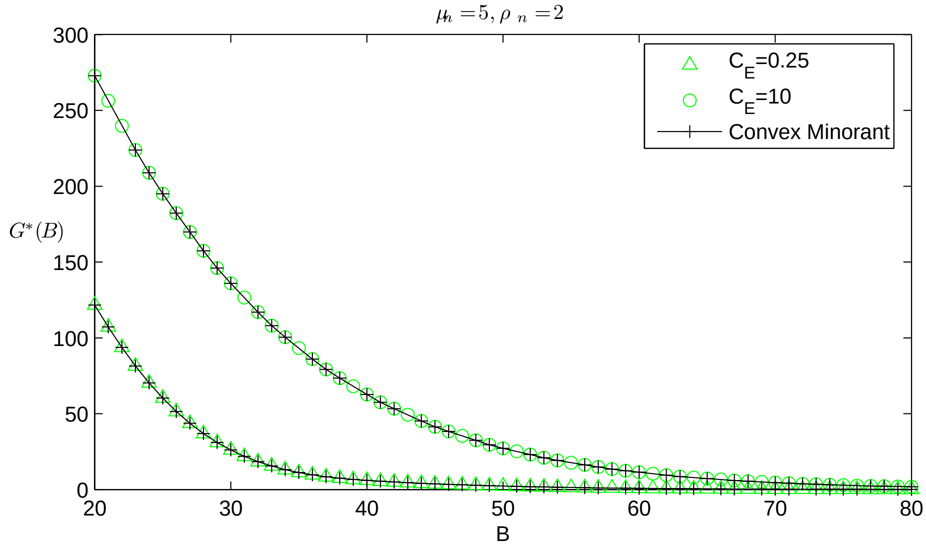

Figure 4.2: Cost Curve with Largest Minimal Convex Minorant for Identical Independent FSLs Case

4.6.3 Inventory Investment Budget Cost Curve

In this section, we explore the shape of the cost function $\tilde{G}^*(B)$ and recommend using a convex minorant of this function for advanced work.

We use bisection search algorithm, (H1) to investigate the shape of cost function $\tilde{G}^*(B)$. Figure 4.2 and 4.3 are the plots of $\tilde{G}^*(B)$ (light color with different shapes) and its corresponding convex minorant (black line with cross), named as $\hat{G}^*(B)$, for the identical and non-identical FSLs. The plots show that $\tilde{G}^*(B)$ is nearly convex and the corresponding largest convex minorant $\hat{G}^*(B)$ provides a very nice approximation of $\tilde{G}^*(B)$.

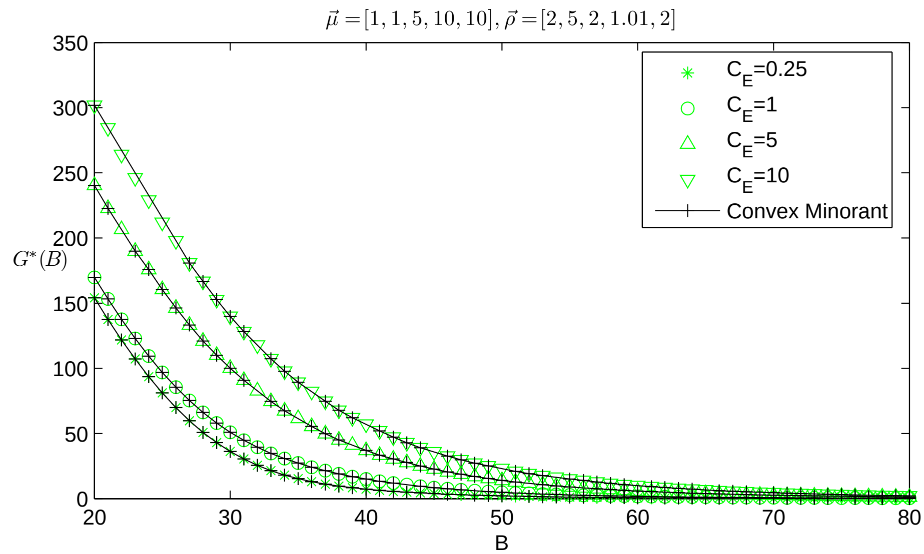

Figure 4.3: Cost Curve with Largest Minimal Convex Minorant for Non-Identical Independent FSLs Case

4.7 Algorithms Optimizing Stock up to Levels for Multiple Items

In this section, we return to the original system with I part types and propose heuristics to search for the optimal solution to problem (4.6). Define an alternative optimization problem as follows:

$$\begin{aligned} \min_{\vec{B}} \quad & \sum_{i=1}^{I} \hat{G}^*(B^{(i)}) \\ \text{s.t.} \quad & \sum_{i=1}^{I} B^{(i)} = B, \\ & B^{(i)} \in Z^+ \text{for} i = 0, 1, \ldots, I. \end{aligned} \tag{4.10}$$

where $\hat{G}^*(B^{(i)})$ is the convex minorant introduced in the previous section.

As discussed in section 4 and subsection 6.3, $\hat{G}^*(B^{(i)})$ provides a good approximation for $G^*(B^{(i)})$. It is anticipated that the optimal solution of problem (4.10) will provide a good approximation for $\vec{B}^*$, the optimal solution of problem (4.6).

We propose the three main elements for the algorithm to solve optimization problem (4.10) as follows:

- **Convex Minorant Local Construction**: For each part type $i = 1, \ldots, I$, we use bisection search algorithm (H2) to compute the value of $\tilde{G}^*(B^{(i)})$ for $B^{(i)} \in [0, U^{(i)}]$, where $[0, U^{(i)}]$ is the region containing the optimal solution $B^{(i)*}$. Initially, $U^{(i)}$ is chosen to be a value much smaller than B.
- **Marginal Analysis Search**: Compute the convex minorant $\hat{G}^*(B^{(i)})$ for $\tilde{G}^*(B^{(i)})$ and use a marginal analysis algorithm to search for the optimal solution of problem (4.10).
- **Upper Bound Update**: If the solution $B^{(i)}$ to the marginal analysis results in $B^{(i)} = U^{(i)}$ for some FSL, then estimate another larger upper bound $U'^{(i)}$ and use bisection search to compute $\tilde{G}^*(B^{(i)})$ for $B^{(i)} \in [U^{(i)}, U'^{(i)}]$. Update $\hat{G}^*(B^{(i)})$ over $[0, U'^{(i)}]$. Continue with the marginal analysis search for problem (4.10).

4.8 Conclusions

We develop optimization algorithms for setting stock levels in a resupply network with both field service locations (FSL) and an emergency stocking location(ESL). We proposed a bisection search algorithm to determine the stock levels at the FSLs and ESL given inventory investment for single item. Since the problem is a problem with a potentially non-convex objective, we use PSWARM as a benchmark to validate the bisection search algorithm.

From the empirical results, we find that when the inventory investment budget is small, as the VTMR of demand at the FSL increases, the optimal solutions incline to stock less at the FSL. While the VTMR is small, the demand rate is the dominating factor in deciding the stock levels at the FSLs. When the budget is small, the ESL plays an important role to stock the inventory shared among the FSLs. However, when the inventory investment budget becomes large enough, the optimal solutions incline to stock more at the FSLs which have higher demand rate and higher variance. Less inventory is kept at the ESL since the emergency orders and its associated costs are mainly reduced by the large amount inventory stocked at the FSLs.

On the other hand, as the emergency penalty cost per order decreases, meaning that the shortage of the FSLs incurs less penalty, the optimal stock levels of the FSLs decrease and it is optimal to stock more at the ESL. This also supports a strategy that when there is little emergency penalty cost, we incline to stock everything at one location, the ESL, to pool the variability of demand at each FSL. Besides, given a fixed inventory investment budget, the expected cost always increases as the mean and the variance of the demand increases.

We conclude that the benefit of the ESL becomes significant when the inventory investment budget is small and the VTMR of the demand is large. Besides, it is recommended to increase the investment budget to control the system cost when the mean and the variance of the demand at the FSLs increase.

After understanding the cost function given any inventory investment budget for single item, we propose the main elements of an algorithm to solve the optimal stock levels for the multi-item and multi-location problem. These elements include using the convex minorant of the cost function for each item and applying marginal analysis over the total inventory investment budget.

BIBLIOGRAPHY

[1] Alfredsson, P. and J. Verrijdt. Modeling Emergency Supply Flexibility in a Two-Echelon Inventory System. *Management Science*, 45 (1999), No. 10, 1416-1431.

[2] Audet, C. and J.E. Dennis. Analysis of generalized pattern searches. *SIAM J. Optim.* 13. 889-903 (2003)

[3] Axsäer, S.. Modeling emergency lateral transshipments in inventory systems. *Management Science*, 36 (1990), 1329-1338.

[4] Axsäter, S. A New Decision Rule for Lateral Transshipments in Inventory Systems, *Management Science*, Vol. 49, No. 9, (Sep. 2003), pp. 1168-1179.

[5] Axsäter, S.. *Inventory Control*, 2nd ed., Springer, New York, 2006.

[6] Axsäter, S.. A heuristic for triggering emergency orders in an inventory system. *European Journal of Operational Research*, 176 (2007) 880C891.

[7] Archibald, T.W., S.A. Sassen, and L.C. Thomas. An optimal policy for a two depot inventory problem with stock transfer. *Management Science*, 43 (1997), 173C183.

[8] Chen, J., P. Jackson, and J. Muckstadt. Exact Analysis of a Lost Sales Model under Stuttering Poisson Demand. Technical Note, *Operations Research* (to appear 2010a). Technical Report No. 1474, School of Operations Research and Information Engineering, Cornell University.

[9] Chen, J., P. Jackson, and J. Muckstadt. An Emergency Resupply Network Model under Stuttering Poisson Demand. (2010b) Submitted. Technical Report No. 1475, School of Operations Research and Information Engineering, Cornell University.

[10] Chiang, C.. Ordering Policies for Inventory Systems with Expediting. *The Journal of the Operational Research Society*, Vol 53 (2002), No.11, 1291-1295.

[11] Cohen, M. A., P. R. Kleindorfer and H. L. Lee, Optimal Stocking for Low Usage Items in Multi-Echelon Inventory Systems, *Naval Research Logistics Quarterly*, 33 (1986), 17-38.

[12] Dada, M., A Two-Echelon Inventory System with Priority Shipments, *Management Science*, 1140–1153 (1992).

[13] Dada, M.. A Two-Echelon Inventory System with Priority Shipments, *Management Science*. 38 (1992), 1140-1153.

[14] Davis, C., Theory of positive linear dependence. *Am. J. Math* 76, 733-746 (1954)

[15] Das, C.. Supply and Redistribution Rules for Two-Location Inventory Systems: One Period Analysis, *Management Science*, 21 (1975), 765-779.

[16] Evers, P.R., Hidden benefits of emergency transshipments. *Journal of Business Logistics.* 18(2):55-77,1997.

[17] Evers, P.R., Filling customer orders from multiple locations: A comparison of pooling methods. *Journal of Business Logistics*, 20(1): 121-140, 1999.

[18] Fox, B.L. and D.M. Landi. Searching for the multiplier in one constraint optimization problems. *Operations Research*, 18(2):253-262, March-April 1970.

[19] Grahovac, J. and A. Chakravarty. Sharing and Lateral Transshipment of Inventory in a Supply Chain with Expensive, Low-Demand Items. *Management Science*, 47 (2001), 579-594.

[20] Graves, S.C. A multi-echelon inventory model for a repairable item with one for one replenishment. *Management Science*, 31(10): 1247-1256, Octorber 1985.

[21] Gross, D., *Centralized Inventory Control in Multilocation Supply Systems* in *Mlltistage In'entor'J Models and Techlniqles*, H. E. Scarf, D. M. Gilford and M. W. Shelly (Eds.), Stanford University Press, Stanford, CA, 1963, 47-84.

[22] Hoadley, B. AND D. P. Heyman, A Two-echelon Inventory Model with Purchases, Dispositions, Shipments, Returns and Transshipments, *Naval Research Logistics Quarterly*, 24 (1977), 1-19.

[23] Johansen, S.G. and A. Thorstenson. An inventory model with Poisson demands and emergency orders. *International Journal of Production Economics*, 56C57 (1998), 275C289.

[24] Karmarkar, U. S., and N. R. Patel, The One-Period, N-Location Distribution Problem,*Naval Research Logistics Quarterly*, 24 (1977), 559-575.

[25] Kranenburg, A. and G. van Houtum . A new partial pooling structure for spare parts networks. *International Journal of Production Economics*, 108 (2009), 221-227.

[26] Kukreja, A., D.M. Miller, and C.P. Schmidt. Stocking decisions for low-usage items in a multilocation inventory system. *Management Science*, 47 (2001), 1371C1383.

[27] Kolda, T.G., R.M. Lewis, and V. Torczon. Optimization by direct search: new perspectives on some classical and modern methods. *SIAM Rev.* 45, 385-482(2003)

[28] Lee, H.L. A multi-echelon inventory model for repairable items with emergency lateral transshipments. *Management Science*, 33(10):1302-1216, Oct 1987.

[29] Minner, S., D.J. Robb, and E.A. Silver, An improved heuristic for deciding on emergency transshipments. European Journal of *Operational Research*, 148 (2003), 384C400.

[30] Moinzadeh, K., and C.P. Schmidt. An (S-1,S) Inventory System with Emergency Orders. *Operations Research*, 39 (1991), No. 2. pp. 308-321.

[31] Muckstadt, J.A. Some approximations in multi item multi echelon inventory systems for recoverable item. *Naval Research Logistics Quarterly*, 25(3),:377-394, September 1978.

[32] Muckstadt, J.. 2005. *Analysis and Algorithms for Service Parts Supply Chains*, Springer.

[33] Muckstadt, J. and A. Sapra. 2010. *Principles of Inventory Management: When You Are Down to Four, Order More*, Springer.

[34] Olsson, F.. Optimal Policies for Inventory Systems with Lateral Transshipments, *Int. J.ProductionEconomics*, 118 (2009), 175-184.

[35] Paterson C., G. Kiesmüller, R. Teunter, and K. Glazebrook, Inventory models with lateral transshipments: A review, Beta Working 2009, Paper 287, Eindhoven University of Technology.

[36] Reijnen, I.C., T. Tan, and G.J. van Houtum. Inventory Planning for spare parts networks with delivery time requirements. Working paper, Eindhoven University of Technology. 2009.

[37] Robinson, L. W.. Optimal and Approximate Policies in Multiperiod, Multilocation Inventory Models with Transshipments. *Operations Research*, 38 (1990), 278-295.

[38] Rosenshine, M. and D. Obee. Analysis of a Standing Order Inventory System with Emergency Orders. *Operations Research*, 24 (1976), 1143-1155.

[39] Rudi, N., S. Kapur, and D. F. Pyke. A Two-Location Inventory Model with Transshipment and Local Decision Making, *Management Science*, 47 (2001), 1668-1680.

[40] Sherbrooke, C.C. Multiechelon Inventory Systems with Lateral Supply. *Naval Research Logistics*, 39 (1992), 29-40.

[41] Sherbrooke, C.C. METRIC A multi echelon technique for recoverable item control. *Operations Research*, 16(1): 122-141, 1968.

[42] Sherbrooke, C.C. VARI-METRIC: Improved approximations for multi indenture, multi-echelon availability models. *Operations Research*, 34(2):311-319, March-April 1986.

[43] Sherbrooke, C.C. Multi-echelon inventory systems with lateral supply. *Naval Research Logistics*, 39:29-40, 1992.

[44] Tagaras, G. and M.A. Cohen, Pooling in two-location inventory systems with non-negligible replenishment lead times. *Management Science*, 38 (1992), 1067C1083.

[45] Tagaras, G. and D. Vlachos. A Periodic Review Inventory System with Emergency Replenishments. *Management Science*, 47 (2001), 415-429.

[46] Teunter, R. and D. Vlachos. An Inventory System with Periodic Regular Review and Flexible Emergency Review. *IIE Transactions*, 33 (2001), 625-635.

[47] Vaz, A.I.F. and L.N. Vicente, A particle swarm pattern search method for bound constrained global optimization, *Journal of Global Optimization*, 39 (2007) 197-219.

[48] Vaz, A.I.F. and L.N. Vicente, PSwarm: A hybrid solver for linearly constrained global derivative-free optimization. Submitted (draft). http://www.norg.uminho.pt/aivaz/pswarm/

[49] Van Den Berghand, F. An analysis of particle swarm optimizers. Ph.D. thesis, Faculty of Natural and Agricultural Science, University of Pretoria (2001).

[50] Vidgren, S.. Evaluation of a Class of Lateral Transshipment Policies, Lund University, 2005.

[51] Vliegen, I.. *Integrated planning for service tools and spare parts for capital goods*, Ipskamp Drukkers, 2009.

[52] Wijk A. van, I. Adan, and G. van Houtum. Optimal lateral transshipment policy for a two location inventory problem. Working paper, Eindhoven University of Technology. 2009.

[53] Wong, H., G.J. Van Houtum, D. Cattrysse, and D. Van Oudheusden, Multi-item spare parts systems with lateral transshipment and waiting time constraints. *European Journal of Operational Research*, 171 (2006), 1071C1093.

[54] Whittmore, A.S. and S. Saunders, Optimal Inventory under Stochastic Demand with Two Supply Options. *SLAM J. Appl. Math.*, 32 (1977), 293-305.

CHAPTER 5

CONCLUSION

This thesis focuses on a system involving multiple field service locations (FSL) and an emergency stocking location (ESL), and considers emergency orders under a stuttering Poisson demand process. First, an exact analysis of the number of units on order is derived for a single FSL by understanding the emergency orders to have the same behavior as lost sales. Later, we build the model for the number of units on order at the ESL by extending the exact results for the single FSL. Finally, we develop an effective algorithm to optimize stock levels at the ESL and the associated cluster of FSLs simultaneously.

There is more research that needs to be conducted in this area. For example, we point out that the exact analysis for the general lost sales distribution is still an open question since our method applies only to the case of stuttering Poisson demand. Furthermore, for more practical use, we would need to fit the stuttering Poisson process to a general demand history and use our results as an approximation. Besides, there are usually thousands of part numbers managed in such system. The calculations for each part number must be extremely efficient and therefore the computation of the matrix inversion in the approximation algorithm, which could become quite costly, needs to be carefully designed. Further approximations may be required for speed-up. In addition, integrating these models with a multi-part budget, as suggested in chapter 4, requires further development and testing.

For more theoretical extensions, we would want to consider the same system with more general demand distributions. We are also interested in a discrete time model. We have observed an interesting shape to the stationary distribution of the number of units in resupply when the demand order size is negative-binomially distributed in the discrete time case. Understanding and approximating this phenomenon will enable us to handle more realistic demand processes. Further research could also be conducted in the direction of more flexible resupply networks such as allowing for inventory to be shared among the ESLs (pooling among clusters).

GLOSSARY

λ rate of customer arrivals

X customer order size

$k = 0, 1, \ldots$ order size

$p_k \equiv P\{X = k\}$

$\overline{P}_k \equiv P\{X > k\}$

I_t inventory on hand at time t

S order up to level

μ delivery rate

$\tau = \frac{1}{\mu}$ expected lead time

N_{kt} the number of replenishment orders size k at time t

$N_t = (N_{kt})_{k=1}^{S}$

$N = \{N_t, t \geq 0\}$

V state space of replenishment orders

$n_k(i)$ number of orders size k, $i \in V$

$n(i) = (n_1(i), n_2(i), ..., n_S(i))$

$n_0(i) \equiv S - \sum_{k=1}^{S} kn_k(i)$

$m(i)$ the number of orders outstanding

(i, j) transition

$\|(i, j)\| \equiv \sum_{k=1}^{S} |n_k(i) - n_k(j)|$

$k_{ij} \equiv \sum_{k=1}^{S} k\,|n_k(i) - n_k(j)|$

V_C^2 customer order arrivals class

V_R^2 replenishment order arrivals class

A_{ij} infinitesimal generator for lost sale Markov process

$\tilde{X}(t)$ generic continuous Markov chain

$\tilde{V}$ generic state space

$\tilde{\xi}_i$ generic stationary distribution

$\tilde{A}$ generic generator for $\tilde{X}$

i_0 reference state

ν_i reversibility rates

$\pi = (\pi_s)$ stationary distribution of number of units on order

$\eta_{m,s} = \sum_{\substack{i \in V \\ S - n_0(i) = s \\ m(i) = m}} \xi_i$ stationary probability of m orders and s units on order

$\bar{\nu} = \frac{1}{1 + \sum_{j \neq i_0} \nu_j}$ normalizing constant

$f_{NB}(\cdot; m, p)$ negative binomial probability distribution

$G(S)$ normalizing constant for (π_s) distribution

$\tau_E = \frac{1}{\mu}$ expected lead time for the emergency case

$Y_k(t)$ the number of emergency orders of size k

Y_k stationary random variable

$Y(t)$ the total number of emergency orders

Y stationary random variable

$Z(t)$ total number of units in emergency resupply

Z stationary random variable

X_t size of customer order at time t

$\tau_F = \frac{1}{\mu_F}$ expected lead time of regular replenishment

$\tau_E = \frac{1}{\mu}$ expected lead time of emergency replenishment

N_{kt} the number of replenishment orders size k at time t

$N_t = (N_{kt})_{k=1}^{S}$

$N = \{N_t, t \geq 0\}$

V state space of replenishment orders

$n_k(i)$ number of orders size k, $i \in V$

$n(i) = (n_1(i), n_2(i), ..., n_S(i))$

$n_0(i) \equiv S - \sum_{k=1}^{S} kn_k(i)$

$m(i)$ the number of orders outstanding

$\pi = (\pi_s)$ stationary distribution of number of units on order

$\xi_{m,s} = \sum_{\substack{i \in V \\ S - n_0(i) = s \\ m(i) = m}} \xi_i$ stationary probability of m orders and s units on order

$f_{NB}(\cdot; m, p)$ negative binomial probability distribution

$G(S)$ normalizing constant for FSL stocking level S

$\cdot^k$ system with emergency orders of size k

y number of emergency order

$N_t^k(i, y) = (N_t = n(i), Y_k(t) = y)$

$\psi_{i,y}^k$ steady state distribution for k^{th} system

$\Lambda_{h,i}^k$ hth moment of outstanding emergency orders of size k with $i \in V$, $h = 1, 2, 3$

$\tilde{\Lambda}_{1,s}^k$ mean of replenishment orders of size k at the ESL with s units on order at the FSL

Λ_h^k hth moment of outstanding emergency orders of size k, $h = 1, 2, 3$

$\rho_k(i)$ arrival rate for $Y_k(t)$ conditioned on $i \in V$

N^+ the approximate Markov process for the number of units in resupply at the FSL

Q infinitesimal generator for approximate Markov chain

$g(m, s, s')$ likelihood of transition to s' given transition from m, s

$\gamma(s, s')$ rate of transitions from s to s'

$\Gamma(s)$ rate of transition's out of state s

$\psi_{y|i}$ steady state distribution of y orders of size k at the ESL conditioned on $i \in V$

$\tilde{\rho}_k(s) = \rho_k(i)$ for any state i with $S - n_0(i) = s$

k_1, k_2 two order sizes at the ESL

$Y_{k_1k_2,t}$ number of emergency orders of size k_1, or k_2 at time t

$\tilde{\rho}_{k_1k_2}(s)$ arrival rate for $Y_{k_1k_2,t}$ given s units outstanding

$\Lambda_{1,i}^{k_1k_2}$ mean of emergency orders of size k_1 or k_2 with i replenishment order vector

$\tilde{\Lambda}_{1,s}^{k_1k_2}$ mean of emergency orders of size k_1 or k_2 and s units on order at the FSL

$\Lambda_1^{k_1k_2}$ mean of emergency orders of size k_1 or k_2

$\Lambda_2^{k_1k_2}$ second moment of emergency orders of size k_1 or k_2

K limit of summation for k

X_L cumulative unit arrivals during time L

σ^2 variance rate of unit arrival process

$\tilde{Z}$ a distribution mixing an atom at zero and a zero-truncated negative binomial distribution